

# Constrained Data Decomposition and Regression for Analyzing Healthy Aging from Fiber Tract Diffusion Properties

Sylvain Gouttard<sup>1</sup>, Marcel Prastawa<sup>1,2</sup>, Elizabeth Bullitt<sup>3</sup>, Weili Lin<sup>4</sup>, Casey Goodlett<sup>1,2</sup>, and Guido Gerig<sup>1,2</sup>

<sup>1</sup> Scientific Computing and Imaging Institute, <sup>2</sup> School of Computing, University of Utah  
Departments of <sup>3</sup> Surgery and <sup>4</sup> Radiology, University of North Carolina  
gouttard@sci.utah.edu \*

**Abstract.** It has been shown that brain structures in normal aging undergo significant changes attributed to neurodevelopmental and neurodegeneration processes as a lifelong, dynamic process. Modeling changes in healthy aging will be necessary to explain differences to neurodegenerative patterns observed in mental illness and neurological disease. Driving application is the analysis of brain white matter properties as a function of age, given a database of diffusion tensor images (DTI) of 86 subjects well-balanced across adulthood. We present a methodology based on constrained PCA (CPCA) for fitting age-related changes of white matter diffusion of fiber tracts. It is shown that CPCA applied to tract functions of diffusion isolates population noise and retains age as a smooth change over time, well represented by the first principal mode. CPCA is therefore applied to a functional data analysis (FDA) problem. Age regression on tract functions reveals a nonlinear trajectory but also age-related changes varying locally along tracts. Four tracts with four different tensor-derived scalar diffusion measures were analyzed, and leave-one-out validation of data compression is shown.

## 1 Introduction

Unlike earlier hypothesis that brain anatomy and major functions are pretty much established after adolescence and would not change significantly until late adulthood, there is increasing evidence of dynamic, lifelong changes of brain structures and thus plasticity. To examine correlation of age-related brain changes, volumetric measurements such as brain tissue [1] or cortical thickness [2] were proposed. Regression on brain MRI has been used by Davis *et al.* [3] to

---

\* This work is part of the National Alliance for Medical Image Computing (NA-MIC), funded by the National Institutes of Health through Grant U54 EB005149. We also acknowledge support from the NIMH Silvio Conte Center for Neuroscience of Mental Disorders MH064065, the BRP grant R01 NS055754-01-02 and R01 EB000219 NIBIB.

depict age-related shape and volume changes. White matter changes in diffusion tensor imaging (DTI) due to aging were reported by analyzing changes of scalar diffusion measures such as fractional anisotropy (FA) and mean diffusivity (MD) in manually selected brain [4,5] or tract regions [6]. Analysis was most often performed by regional measurements of FA or MD across different age groups, or by linear regression of such measurements versus age. Whereas most of this previous analysis uses manually selected regions of interest strategically placed in subregions of known fiber tract locations, clinical research is often interested in the analysis of *whole fiber tracts* associated with specific tasks or cognitive function. Age-related changes of tract diffusion properties should therefore be represented at various positions of tracts, informing researchers about anatomical location and type of diffusion changes. This paper presents a methodology for age regression of fiber tract diffusion properties. We follow the work of Goodlett *et al.*[7] where fiber bundles are parametrized with arc-length and attributed with local diffusion properties summarized across perpendicular cross-sections. These parametrized representations with diffusion attributes can thus be represented by functions whose shape represents variation of diffusion as a function of locality,  $f(s)$ , with  $s = [0 \dots l]$ . Diffusion is a scalar derived from tensors, so that  $f(s)$  can be a function of FA or MD, for example. Measuring such tract functions from subjects with different age  $t$  results in a set of functions parametrized as  $f(t, s)$ . Using a constrained PCA technique (CPCA), we demonstrate how the complex shape and time change information encoded in  $f(t, s)$  can be simplified to a model where regression on one coefficient efficiently represents locality, type and magnitude of age-related diffusion changes.

This paper is organized as follows: Section 2 describes the concept of constrained PCA. The processing pipelines for population-based segmentation of fiber tracts and for analysis of age-related diffusion changes of tracts are described in section 3. Section 4 summarizes results on several tracts from the healthy aging study, followed by validation experiments (section 5).

## 2 Constrained Principal Component Analysis (CPCA)

Our problem is defined as follows: Given a set of parametrized functions of diffusivity attributed with age,  $f(t, s)$ , we would like to isolate the major systematic change of functional shapes with respect to age from population noise. Such a dimensionality reduction cannot be optimally achieved by principal component analysis (PCA) since any decomposition to modes of major variability would not consider the time attribute. We therefore fit the set of functions into a space that is constrained to a smooth change with time. The first or first few of the resulting coefficients can then be used for age regression via polynomial fit.

We propose the use of constrained Principal Component Analysis (CPCA) [8,9] to perform data reduction that incorporates the age or time effect in the observed data. For a set of discretized functions  $f(t, s)$  that represents the observed population data at different points of time, we first construct a data matrix  $Z_{t,s}$  indexed by time  $t$  along the row and indexed by the tract spatial

parametrization  $s$  along the column (as in standard PCA). CPCA linearly separates the data matrix given multiple constraint variables or external information about the structure of the data along the row or columns of the observations. For constraint variables  $\alpha_1, \dots, \alpha_j$ , a data matrix  $M$  is rewritten as:

$$M = \tilde{M}(\alpha_1) + \dots + \tilde{M}(\alpha_j) + R \quad (1)$$

where  $R$  is the residual data matrix. Unlike standard PCA, CPCA decomposes the individual data matrices for each constraint variable rather than the original unconstrained data matrix. This ensures that the subspaces generated by CPCA encode the effects of the constraint variables. In our tract analysis, we have one constraint variable along the row (time) and  $Z$  can be formulated as:

$$Z = \tilde{Z}(t) + R \quad (2)$$

where  $\tilde{Z}(t)$  is a projection of each individual function onto the space of polynomials  $a_0 + a_1t + \dots + a_k t^k$  that smoothly varies with time and  $R$  is the residual data matrix. We have chosen the projection to the space of polynomials to constrain the analysis to changes that vary smoothly with time and exclude noise and population variability. Given a matrix of polynomial basis functions  $B$  and the matrix of polynomial coefficients that best fit the observed data  $C$ ,  $\tilde{Z}(t) = BC$  and  $Z = BC + R$ .

Within the constrained CPCA framework, all analysis of the diffusion properties is performed within the subspaces of the constrained data matrix  $BC$  and we exclude the residual matrix to ensure that noise is not attributed as age effect. A polynomial up to a degree 4 was used in our experiments.

### 3 Application to age-related white matter tract changes

*Driving application: White matter change in healthy aging:* We have access to a database of high-resolution, 3-Tesla MR images of 100 healthy subjects aged 20 to 76, with 20 subjects per decade divided equally by gender. All subjects are carefully screened for the presence of brain-related disease. Images acquired include T1, T2, MRA, and diffusion tensor images. After screening for image quality of successful DTI scans, 86 DTI could be selected for this study. DTI was done on a Siemens Allegra 3T head-only scanner: One image without diffusion gradients together with diffusion weighted images along six gradient directions with a b-value of 850, repeated 2 times for averaging. TR/TE were 7500/73ms, the matrix size was 64x64, and voxel resolution was isotropic with  $2x2x2mm^3$ .

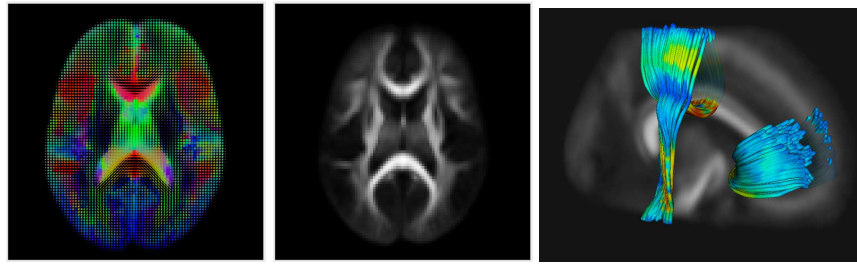
*Atlas building of tensor images:* The images are initially aligned by affine registration of the non-diffusion weighted images to a template T2 (MNI) image using mutual information, which ensures mapping into a normative brain space. A scalar feature, the greatest eigenvalue of the Hessian of the FA image, is computed for each subject [7]. A diffeomorphic, fluid registration is computed for the set of feature images [10] using the affine alignment as an initialization. The

result of the registration is a set of transformations that smoothly map each subject image into a common atlas coordinate system and can also be used to map structures from the atlas back to each subject's native space. All tensor images deformed into the atlas space are averaged to produce a template tensor image for the atlas, which is used to extract fiber tracts for the white matter structures of interest. We extracted the four bundles genu, splenium, mid corpus callosum (mid-cc) and motor-sensory via fiber tractography in atlas space to ensure correspondence across subjects (Fig. 1 right), using a standard streamline integration method. These tracts were then mapped back into each subjects native space to obtain diffusion measures for that subject. This results in a set of tracts in the atlas space with equivalent geometry but diffusion values extracted from each subject. Validation and comparison to individual tractography has been done, but space constraints prohibit inclusion in this paper. Since subjects have different ages equally distributed between 19 and 68, we therefore get sets of functions represented as  $f(t, s)$  with  $s$  representing parametrization by arc-length and  $t$  subject age. The scalar diffusion measures  $f$  can be chosen as FA, MD (trace), axial diffusivity (AD,  $\lambda_1$ ) and radial diffusivity (RD,  $(\lambda_2 + \lambda_3)/2$ ).

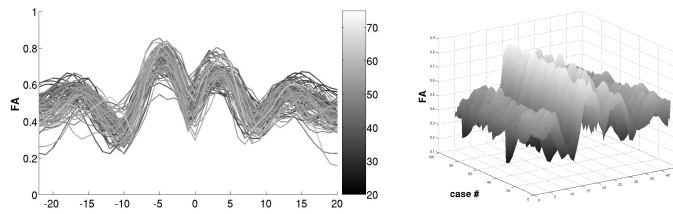
*CPCA on a set of functions:* Figure 2 illustrates the example of FA of the mid-cc tract represented by a set of functions coded with age. We can observe that FA at the middle part close to the corpus callosum does not change strongly with age but that the sidelobes, i.e. bundle locations close to the cortex, show significant decrease of FA with age. The complexity of age change and the type and amount of change at different tract locations are not visible due to large population variability. We then applied constrained CPCA to the set  $FA(t, s)$  in order to decrease dimensionality and determine major shape variation of the FA functions with age. Standard PCA is applied for comparison.

Figure 3 illustrates results for PCA (top) and CPCA (bottom) for the mid-cc tract. Plots in the left column show samples projected into the space of the first two principal components and age. Samples are color-coded with age. As expected, PCA does not present obvious structure versus age but CPCA results in a low-dimensional continuous path for the 86 samples. The middle column illustrates projection to only the first component versus age, with overlay of 4th-order regression on age. Again, PCA does not depict a clear relationship to age, but age regression of CPCA (middle, bottom) illustrates a nonlinear continuous change with age, with large changes between age 20 and 30, than flattening till age 50, and accelerated change thereafter. The right column displays the FA tract functions reconstructed from the first component for all 86 subjects. Unlike PCA, CPCA encodes age change of the FA functions, clearly visible as the systematic age-sorted coloring of CPCA and random coloring of PCA.

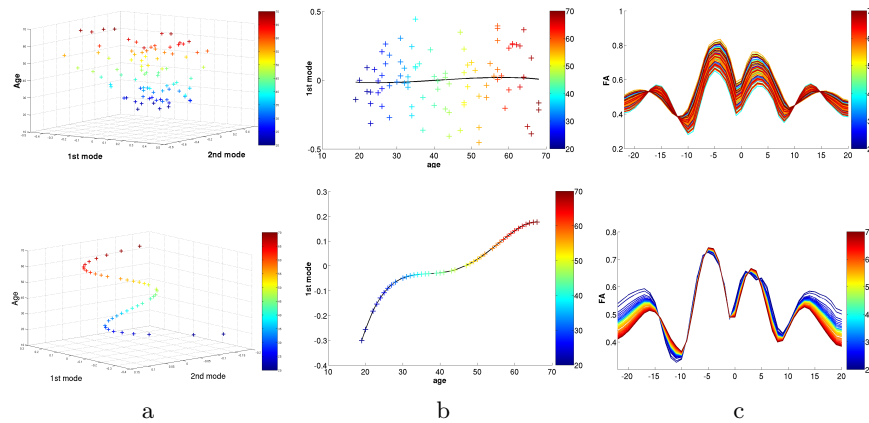
Our main aim is a simplified model of changes of diffusion functions with respect to age. As seen in Fig. 4(b) left for the mid-cc tract, most energy is concentrated on the first component of CPCA. Moreover, age correlation is very strong for the first component but nonsignificant for the others. PCA is shown for comparison. We therefore chose to model the diffusion changes by age regression within the subspace of only the first component. Here, we fit a polynomial up to



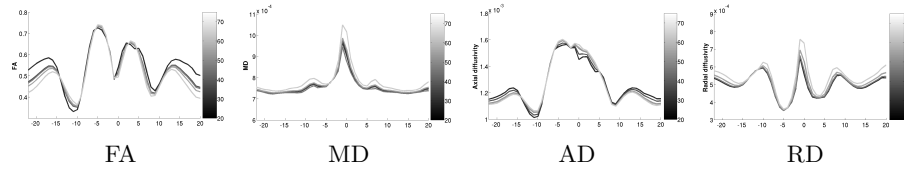
**Fig. 1.** DTI population atlas: Left Tensor field of atlas computed over 86 subjects; Middle: FA map calculated from tensor atlas. Right: Typical fiber bundles used in this study. Mid corpus callosum, sensory-motor and genu tracts are color-coded with FA, with  $[0 \cdots 1]$  ranging from dark blue over yellow to red.



**Fig. 2.** Plot of all mid-cc tract functions of FA shaded by age (left) and 3D representation of the age set sorted by age (young to old from right to left).



**Fig. 3.** Results of the analysis of FA of the mid-cc tract with standard PCA (top) and constrained CPCA (bottom). Age is encoded in color. a) The first two modes plotted against age. b) Only first mode versus age with overlay of 4th order age regression. c) FA functions of 86 subjects projected onto the first component. Columns b and c demonstrate that data reduction by CPCA does encode a strong age-related relationship not seen in PCA (see random coloring in PCA versus age-sorted coloring in CPCA in column c).



(a) Reconstruction of diffusion properties of the mid-cc tract using polynomial age fitting within the first subspace generated by CPCA. Age is encoded in different shades of gray and shown for years 20,30,40,50 and 60. We display the evolution of diffusion properties as a function of age for FA, MD, AD and RD.

Mode	CPCA		PCA		Mode	CPCA %eval			
	%eval	corr	%eval	corr		FA	MD	AD	RD
1	66.20	<b>0.978</b>	24.25	0.069	1	66.20	80.98	55.40	64.91
2	23.68	0.032	20.82	-0.496	2	23.68	15.91	21.92	25.93
3	7.78	0.020	12.86	0.145	3	7.78	2.54	19.70	6.87
4	2.33	0.203	10.81	-0.047	4	2.33	0.56	2.96	2.27

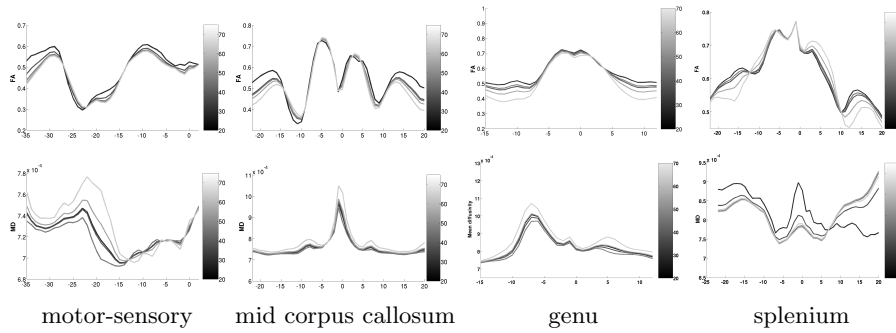
(b) Left: Normalized eigenvalues and age correlation of first four CPCA and PCA modes on the set of functions of FA along the mid-cc tract, illustrating the stronger compression of CPCA but also concentration of age-correlation on the first component. Right: Normalized eigenvalues of CPCA for FA, MD, AD and RD diffusion measures.

**Fig. 4.** Results of mid-cc tract for four different diffusion measures.

a degree 4 to the projections of the functions to the first component (see Fig. 3 column b). Using this polynomial fit, we can then reconstruct the expected diffusion properties for each age in the range of interest. Fig. 4(a) illustrates the reconstructions for ages 20,30,40,50 and 60 and for various diffusion properties (FA, MD, AD, RD) along the mid-cc tract. Age is encoded using a gray-scale color map. The figures show nonlinear differences across age and also variations of change patterns as a function of tract location.

## 4 Results

The CPCA analysis has been applied to the major four fiber tracts genu, splenium, mid-cc and motor-sensory. Fig. 5 displays age-related changes as calculated for these four tracts. We illustrate the reconstructed diffusion functions of FA and MD for age 20, 30, 40, 50 and 60, using projection onto the first CPCA component and 4th-order regression. The plots clearly illustrate that the age-related trajectory of white matter diffusion is nonlinear, i.e. different decades present different changes, and second that diffusion changes vary significantly as a function of tract location. FA in genu, splenium and mid corpus callosum, for example, shows only minor age change in the center region but significant FA decrease in the peripheral parts close to the cortex. The motor-sensory tract (arc-length left to right presents superior to inferior tract location, see also Fig. 1) also demonstrates age-changes of FA close to the cortex but large MD changes



**Fig. 5.** Reconstructed functions of FA (top row) and MD (bottom row) for four fiber tracts. Age is encoded as gray-scale for years 20, 30, 40, 50, and 60. Age-correlation of CPCA was significant for all 4 tracts (0.93, 0.98, 0.85 and 0.95).

along the middle part between cortex and internal capsule. The splenium tract showed much more population variability than the genu, which is reflected in the MD analysis. Major conclusions related to the specific age study might be too preliminary, but this discussion demonstrates how the results could be interpreted by clinical researchers. The caption in Fig. 5 also lists the age correlation for the first component of CPCA which is highly significant for all 4 tracts.

## 5 Validation

We validate the choice of an age-constrained subspace by leave-one-out cross validation. Given a set of  $N$  functions, we perform leave-one-out analysis by excluding one function from the set followed by performing data reduction and regression using the reduced set. We then project the one excluded function to the computed first principal component, and compare the projected score against the expected score from the polynomial fitting. We decided to use this approach to validate the data reduction by projection to a lower dimensional subspace and also the age-based fitting within this subspace. This analysis was performed on the subspaces computed by CPCA and regular PCA. The values for the average of the  $\sum_x |proj(x) - fit(x)|^2$  measure for motor-sensory, mid-cc, genu and splenium are (0.01975, 0.02298, 0.01698, 0.03778) for CPCA and (0.02499, 0.03051, 0.02367, 0.12504) for PCA. As seen, the CPCA consistently provides lower prediction error compared to the unconstrained PCA. The prediction error for PCA is higher for the noisy diffusion data in the splenium tract.

## 6 Discussion

We present a methodology for analyzing age-related changes of white matter measured along fiber tracts. The framework includes population-based mapping of DTI to a common coordinate space, tractography, and representation of

tensor-derived scalar diffusion measures along tracts. CPCA constrained to age is then applied to the sets of diffusion functions to reduce dimensionality to the first principal mode. The PCA projections are performed on the curves reconstructed from the polynomials. We thus enforce the dominant eigenvectors to be strongly driven by time and not necessarily be the original eigenvectors. We preferred this global approach versus regression at each spatial position to account for along-tract correlation. Results on major tracts demonstrate feasibility and represent type of changes of FA and MD as reported elsewhere. However, our preliminary analysis predicts that age changes might not be simplified to linear regression but show more complex time and space variations. The strong dimensionality reduction while still encoding age-related changes as a function of tract location might be a benefit for biostatistical analysis. We plan to apply the method to clinical studies of aging and early brain development, with testing for group differences of change trajectories of white matter diffusion. A limitation might be the independent analysis of tensor-derived diffusion parameters. In the future, we will explore joint analysis of multiple features, e.g. FA and tensor norm or AD and RD, or even full tensors.

## References

1. Mortamet, B., Zeng, D., Gerig, G., Prastawa, M., Bullitt, E.: Effects of Healthy Aging Measured By Intracranial Compartment Volumes Using a Designed MR Brain Database. *Lecture Notes in Computer Science* **3749** (2005) 383–391
2. Sowell, E.R., Thompson, P.M., Leonard, C.M., Welcome, S.E., Kan, E., Toga, A.W.: Longitudinal mapping of cortical thickness and brain growth in normal children. *J. Neurosci.* **24** (2004) 8223–8231
3. Davis, B., Fletcher, P., Bullitt, E., Joshi, S.: Population Shape Regression From Random Design Data. In: *Int. Conf. Comp. Vis., ICCV 2007.* (2007) 1–7
4. Salat, D., Tuch, D., Greve, D., van der Kouwe, A., Hevelone, N., Zaleta, A., Rosen, B., Fischl, B., Corkin, S., Rosas, H., et al.: Age-related alterations in white matter microstructure measured by diffusion tensor imaging. *Neurobiology of Aging* **26** (2005) 1215–1227
5. Pfefferbaum, A., Sullivan, E.: Increased brain white matter diffusivity in normal adult aging: Relationship to anisotropy and partial voluming. *Magnetic Resonance in Medicine* **49** (2003) 953–961
6. Bonekamp, D., Nagae, L., Degaonkar, M., Matson, M., Abdalla, W., Barker, P., Mori, S., Horská, A.: Diffusion tensor imaging in children and adolescents: Reproducibility, hemispheric, and age-related differences. *Neuroimage* **34** (2007) 733–742
7. Goodlett, C.B., Fletcher, P.T., Gilmore, J.H., Gerig, G.: Group analysis of DTI fiber tract statistics with application to neurodevelopment. *NeuroImage* **45** (2009) S133 – S142 *Mathematics in Brain Imaging.*
8. Takane, Y., Hunter, M.A.: Constrained principal component analysis: a comprehensive theory. *Applicable Algebra in Engineering, Communication, and Computing* **12** (2001) 391–419
9. Hunter, M.A., Takane, Y.: Constrained principal component analysis: various applications. *Journal of Educational and Behavioral Statistics* **27** (2002) 41–81
10. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. In: *NeuroImage. Volume 23.* (2004) 151–160