



UNIVERSITY OF LEEDS

This is a repository copy of *Robust constrained weighted least squares for in vivo human cardiac diffusion kurtosis imaging*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/230044/>

Version: Accepted Version

---

**Article:**

Coveney, S. [orcid.org/0000-0002-7134-3196](https://orcid.org/0000-0002-7134-3196), Afzali Deligani, M., Mueller, L. et al. (4 more authors) (Accepted: 2025) Robust constrained weighted least squares for in vivo human cardiac diffusion kurtosis imaging. *Magnetic Resonance in Medicine*. ISSN 0740-3194 (In Press)

<https://doi.org/10.1002/mrm.70037>

---

This is an author produced version of an article accepted for publication in *Magnetic Resonance in Medicine*, made available under the terms of the Creative Commons Attribution License (CC-BY), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

# Robust constrained weighted least squares for in vivo human cardiac diffusion kurtosis imaging

Sam Coveney, Maryam Afzali, Lars Mueller,  
Irvin Teh, Filip Szczepankiewicz,  
Derek K. Jones, Jurgen E. Schneider

Accepted\* 29 July 2025

## Abstract

**Purpose** Cardiac diffusion tensor imaging (cDTI) can investigate the microstructure of heart tissue. At sufficiently high b-values, additional information on microstructure can be observed, but the data require a representation such as diffusion kurtosis imaging (DKI). cDTI is prone to image corruption, which is usually treated with shot-rejection but which can be handled more generally with robust estimation. Unconstrained fitting allows DKI parameters to violate necessary constraints on signal behaviour, causing errors in diffusion and kurtosis measures. **Methods** We developed robust constrained weighted least squares (RCWLS) specifically for DKI. Using in vivo cardiac DKI data from 11 healthy volunteers collected with a Connectom scanner up to b-value  $1350 \text{ s/mm}^2$ , we compared fitting techniques with/without robustness and with/without constraints. **Results** Constraints, but not robustness, made a significant difference on all measures. Robust fitting corrected large errors for some subjects. RCWLS was the only technique that showed radial kurtosis to be larger than axial kurtosis for all subjects, which is expected in myocardium due to increased restrictions to diffusion perpendicular to the primary myocyte direction. For  $b = 1350 \text{ s/mm}^2$ , RCWLS gave the following measures across subjects: mean diffusivity (MD)  $1.68 \pm 0.050 \times 10^{-3} \text{ mm}^2/\text{s}$ , fractional anisotropy (FA)  $0.30 \pm 0.013$ , mean kurtosis (MK)  $0.36 \pm 0.027$ , axial kurtosis (AK)  $0.26 \pm 0.027$ , radial kurtosis (RK)  $0.42 \pm 0.040$ , and RK/AK  $1.65 \pm 0.19$ . **Conclusion** Fitting techniques utilizing both robust estimation and convexity constraints, such as RCWLS, are essential to obtain robust and feasible diffusion and kurtosis measures from in vivo cardiac DKI.

## 1 Introduction

Cardiac diffusion weighted imaging (cDWI) is a magnetic resonance imaging (MRI) technique that can be used to investigate cardiac tissue microstructure. Cardiac diffusion tensor imaging (cDTI) is the most common cDWI method used on the heart, from which measures such as mean diffusivity (MD) and fractional anisotropy (FA) can be derived. However, the diffusion-weighted signal in tissue deviates from monoexponential decay at higher diffusion weighting (as expressed by the b-values) due to cell membranes and other restrictions in biological tissue [1, 2, 3]. Diffusion kurtosis imaging (DKI) can quantify these deviations. Non-Gaussian diffusion models (including DKI) have been shown to have a higher sensitivity for the detection of hypertrophy in ex vivo rat hearts compared with DTI [3]. Kurtosis measures include mean kurtosis (average kurtosis across all directions), axial kurtosis (kurtosis in the primary diffusion direction) and radial kurtosis (average kurtosis in the plane perpendicular to axial kurtosis) [4, 5, 6, 7, 8, 9, 10, 11, 12]. Anisotropic and isotropic kurtosis can be also distinguished with q-space trajectory imaging [13].

---

\*Published article: <https://doi.org/10.1002/mrm.70037>

Spin echo-based DKI in the human heart in vivo is challenging due to low SNR, a short myocardial T2 (approximately 46 ms at 3.0 T [14]) and long echo-times (TE), required to achieve sufficient motion compensation and b-values. Nonetheless, acquiring data with sufficiently high b-values for cardiac DKI has been shown to be feasible in healthy volunteers in vivo using ultra-strong gradients (i.e. 300 mT/m) at echo times and resolutions comparable to those commonly used for conventional cDTI [15, 16, 17]. However, even for the brain, which has longer T2 and significantly less motion, DKI is challenging: fitting methods need to handle data corruptions [18], but also need to yield a physically plausible signal if kurtosis measures are to be meaningful [19].

Image corruption is a common problem in cDWI [20]. Furthermore, motion causes additional signal variations sometimes referred to as physiological noise. In cDTI, shot-rejection is usually performed in an attempt to handle corruptions where noticeably corrupted images are removed from datasets before fitting, a method that is typically time-consuming and subjective. In our experience, the reduced signal at higher b-values simultaneously causes a larger number of corruptions (including those from misregistration) and a decreased ability to perform shot-rejection effectively. Robust estimation, in which outlier signals are identified and removed at the voxel level, is an alternative to shot-rejection. Our recent work shows that robust estimation is superior to shot-rejection in cDTI [21], so robust estimation in cardiac DKI is worth investigating, but has not been done yet.

Although the DKI signal representation does not correspond to a valid diffusion propagator, the fitted signal should still adhere to the physical principles governing the data-generating process. If it does not, the fitted parameters and any measures derived from them will lack meaningful interpretation. For example, the compartment model predicts that kurtosis should be non-negative [2, 4], and the diffusion tensor should be positive definite as in DTI. For DKI, it is not known how to enforce constraints on kurtosis via reparameterization of the fitting problem, so constraints must consider whether the predicted signal behavior is valid, e.g., in [22], non-linear optimization is (infinitely) penalized if constraints are violated. Recently, linear least squares methods have been developed for enforcing convexity constraints on the cumulant generating function, leading to correction of significant errors in brain DKI [19]. These advantages should also apply to cardiac DKI.

While preventing estimated parameters from violating constraints may be seen as a form of robustness, constrained fitting itself is not inherently robust. This is because the constraints do not change the shape of the fitting cost function in parameter space, which is entirely determined by the data. As a result, outliers can still have a detrimental impact on parameter estimates. Furthermore, although robust fitting may reduce the frequency and impact of constraint violations by removing outlier signals, it cannot guarantee that the parameters that optimize the cost function will not violate the constraints. The objectives of this work are: (1) to demonstrate a way of combining robust fitting [21] with convexity constraints [19] using iteratively reweighted least squares (IRLS) to give robust constrained weighted least squares (RCWLS); (2) to test various fitting methods (with/without robustness and with/without constraints) on in vivo cardiac DKI data collected on a Connectom scanner to determine the effects on diffusion and kurtosis measures. To our knowledge this is the first time that constrained estimation has been combined with robust estimation in MRI.

## 2 Methods

We use the following notation: tensors are bold and upper-case; vectors are bold and lower-case; tensor and vector elements are italicized and indexed.

## 2.1 Diffusion Kurtosis Imaging

The DKI signal representation can be expressed as [2]:

$$\begin{aligned} \ln S(\mathbf{q}) = \ln S_0 - \sum_{i=1}^3 \sum_{j=1}^3 q_i q_j D_{ij} \\ + \frac{1}{6} \bar{D}^2 \sum_{i=1}^3 \sum_{j=1}^3 \sum_{k=1}^3 \sum_{l=1}^3 q_i q_j q_k q_l W_{ijkl} \end{aligned} \quad (1)$$

where  $\mathbf{q} = \sqrt{b} \cdot (n_1, n_2, n_3)$  is the (rescaled) wave-vector (denoted this way for convenience in expressing the constraints - see Sec 2.3), and  $i, j, k, l$  index physical space coordinates. The diffusion tensor  $\mathbf{D}$  and kurtosis tensor  $\mathbf{W}$  are both symmetric, having 6 and 15 unique elements, respectively. The signal at  $b = 0$  s/mm<sup>2</sup> is denoted by scalar quantity  $S_0$ . The DTI signal representation is the same as Eq (1) but without the term containing  $\mathbf{W}$ . Kurtosis is expressed in dimensionless form due to scaling by the mean diffusivity  $\bar{D} = (D_{11} + D_{22} + D_{33}) / 3$ .

Expanding Eq (1) accounting for the symmetry of  $\mathbf{D}$  and  $\mathbf{W}$  gives the following linear expression:

$$f_{\theta}(\mathbf{q}) \equiv \ln S(\mathbf{q}) = \boldsymbol{\theta}^T \cdot \mathbf{x} \quad (2a)$$

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_{22})^T \quad (2b)$$

$$\begin{aligned} \mathbf{x} = & (-q_1^2, -2q_1q_2, -q_2^2, -2q_1q_3, -2q_2q_3, -q_3^2, \\ & \frac{1}{6}q_1^4, \frac{1}{6}q_2^4, \frac{1}{6}q_3^4, \frac{4}{6}q_1^3q_2, \frac{4}{6}q_1^3q_3, \frac{4}{6}q_2^3q_1, \frac{4}{6}q_2^3q_3, \frac{4}{6}q_3^3q_1, \frac{4}{6}q_3^3q_2, \\ & q_1^2q_2^2, q_1^2q_3^2, q_2^2q_3^2, 2q_1^2q_2q_3, 2q_2^2q_1q_3, 2q_3^2q_1q_2, -1)^T \end{aligned} \quad (2c)$$

The 22 coefficients in  $\boldsymbol{\theta}$  are related to the original parameters as follows: 6 diffusion tensor parameters  $(\theta_1, \dots, \theta_6) = (D_{11}, D_{12}, D_{22}, D_{13}, D_{23}, D_{33})$ , 15 kurtosis tensor parameters  $(\theta_7, \dots, \theta_{21}) = \bar{D}^2 \cdot (W_{1111}, W_{2222}, W_{3333}, W_{1112}, W_{1113}, W_{1222}, W_{2223}, W_{1333}, W_{2333}, W_{1122}, W_{1133}, W_{2233}, W_{1123}, W_{1223}, W_{1233})$ , and intercept  $\theta_{22} = -\ln S_0$ . The DTI expression would include only terms depending on  $(\theta_1, \dots, \theta_6)$  and  $\theta_{22}$ .

## 2.2 Weighted Least Squares

Given  $N$  observations  $\{\mathbf{q}_n, S_n | n = 1 \dots N\}$ , the weighted least squares (WLS) estimate of the coefficients  $\boldsymbol{\theta}$  in equation (2), denoted by  $\hat{\boldsymbol{\theta}}_{\text{WLS}}$ , is given by:

$$\hat{\boldsymbol{\theta}}_{\text{WLS}} = \arg \min_{\boldsymbol{\theta}} \sum_{n=1}^N w_n \left( f_{\theta}(\mathbf{q}_n) - \ln S_n \right)^2 \quad (3)$$

The wavevectors  $\mathbf{q}_n$  of the  $n$ th observation can be converted to  $\mathbf{x}_n$  using equation (2c). Given design matrix  $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)^T$ , observation vector  $\mathbf{y} = (\ln S_1, \ln S_2, \dots, \ln S_N)^T$ , and weights vector  $\mathbf{w} = (w_1, w_2, \dots, w_N)^T$ , a weighted design matrix and observation vector can be defined:

$$\mathbf{X}' = \text{diag}(\sqrt{\mathbf{w}}) \cdot \mathbf{X}, \quad \mathbf{y}' = \text{diag}(\sqrt{\mathbf{w}}) \cdot \mathbf{y} \quad (4)$$

The WLS estimate can then be written as:

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{\text{WLS}} = \arg \min_{\boldsymbol{\theta}} ||\mathbf{X}' \cdot \boldsymbol{\theta} - \mathbf{y}'||^2 \\ = (\mathbf{X}'^T \cdot \mathbf{X}')^{-1} \cdot \mathbf{X}'^T \cdot \mathbf{y}' \end{aligned} \quad (5)$$

For uniform weights  $w_n = 1$ , Eq (5) gives the ordinary least squares (OLS) estimate  $\hat{\boldsymbol{\theta}}_{\text{OLS}}$ .

### 2.3 Convexity Constraints

A useful constraint for DKI is to enforce convexity of the cumulant generating function  $\mathcal{C}(\mathbf{q})$  [19]:

$$\mathcal{C}(\mathbf{q}) := \ln S(\sqrt{-1}\mathbf{q}) \equiv f_{\boldsymbol{\theta}}(\sqrt{-1}\mathbf{q}) \quad (6)$$

This constraint can be enforced by using sum of squares polynomials [19]; the mathematical background for this can be found in [23]. The semi-definite program for solving the WLS problem subject to constraints, thus yielding the constrained WLS estimate  $\hat{\boldsymbol{\theta}}_{\text{CWLS}}$ , can be written as follows:

$$\hat{\boldsymbol{\theta}}_{\text{CWLS}} = \arg \min_{\boldsymbol{\theta}, \boldsymbol{\alpha}} \|\mathbf{X}' \cdot \boldsymbol{\theta} - \mathbf{y}'\|^2 \quad (7a)$$

$$\text{subject to } \mathbf{e}^T \cdot \left( H(\boldsymbol{\theta}) + L(\boldsymbol{\alpha}) \right) \cdot \mathbf{e} = h_{\boldsymbol{\theta}}(\mathbf{q}, \mathbf{s}) \quad (7b)$$

$$H(\boldsymbol{\theta}) + L(\boldsymbol{\alpha}) \succcurlyeq 0 \quad (7c)$$

$$\text{where } h_{\boldsymbol{\theta}}(\mathbf{q}, \mathbf{s}) := \mathbf{s}^T \cdot H_{\mathcal{C}}(\mathbf{q}) \cdot \mathbf{s} \quad (7d)$$

$$\mathbf{e}^T \cdot H(\boldsymbol{\theta}) \cdot \mathbf{e} = h_{\boldsymbol{\theta}}(\mathbf{q}, \mathbf{s}) \quad (7e)$$

$$\mathbf{e}^T \cdot L(\boldsymbol{\alpha}) \cdot \mathbf{e} = 0 \quad (7f)$$

Importantly, we have written the problem in terms of the weighted design matrix  $\mathbf{X}'$  and observation vector  $\mathbf{y}'$ .

We will briefly explain Eq (7) for DKI, leaving details on the constraint matrices  $H(\boldsymbol{\theta})$  and  $L(\boldsymbol{\alpha})$  (presented here for the first time) for Appendix B. Eq (7d) defines  $h_{\boldsymbol{\theta}}(\mathbf{q}, \mathbf{s})$ , where  $H_{\mathcal{C}}(\mathbf{q})$  is the Jacobian of  $\mathcal{C}(\mathbf{q})$  and the dummy variable  $\mathbf{s}$  has the same dimensions as  $\mathbf{q}$ . Then,  $h_{\boldsymbol{\theta}}(\mathbf{q}, \mathbf{s})$  is just a polynomial. Convexity of  $\mathcal{C}(\mathbf{q})$  requires that  $h_{\boldsymbol{\theta}}(\mathbf{q}, \mathbf{s})$  is non-negative, which can be enforced using a sum of squares polynomial representation, i.e. Eq (7b) (see [23]). We can *exactly* represent  $h_{\boldsymbol{\theta}}(\mathbf{q}, \mathbf{s})$  using a relatively small monomial basis for  $\mathbf{e}$  (see Appendix B). The convexity constraint is satisfied when Eq (7c) holds i.e. when  $H(\boldsymbol{\theta}) + L(\boldsymbol{\alpha})$  is positive semi-definite (PSD). Note that Eq (7) involves optimizing over coefficients  $\boldsymbol{\theta}$  and slack parameters  $\boldsymbol{\alpha}$ . Appendix C explains how the numerical complexity of the problem can be reduced.

### 2.4 Robust Fitting

In DTI/DKI, WLS usually refers to solving Eq (5) by weighting the (squared) residuals of the linearized problem with the (squared) signal [24]:

$$w_n = \left( \exp f_{\boldsymbol{\theta}}(\mathbf{q}_n) \right)^2 \leftarrow \hat{\boldsymbol{\theta}}_{\text{OLS}} \quad (8)$$

where an initial OLS estimate  $\hat{\boldsymbol{\theta}}_{\text{OLS}}$  is used to predict the signals (since the true signals are not known). Henceforth, we will use “WLS” to refer to Eq (5) using the weights in Eq (8). Correspondingly, “CWLS” will refer to constrained WLS, i.e. solving Eq (7) using the weights given by Eq (8).

Importantly, neither WLS or CWLS are intrinsically robust, and outlier data can have a detrimental effect on the estimates  $\hat{\boldsymbol{\theta}}_{\text{WLS}}$  or  $\hat{\boldsymbol{\theta}}_{\text{CWLS}}$ . Robust estimation can be implemented using Iteratively Reweighted Least Squares (IRLS), using weights derived from a robust estimator [25, 18, 21]. Notably, for WLS in DTI/DKI, these robust weight should be chosen so as to preserve the cost function implied by equations (3) and (8) (see [21] for a derivation), such that IRLS solves the WLS/CWLS problem in a robust way. Robust fitting in the DTI/DKI literature is usually done in order to remove the influence of outlier data on the fitted signal, thus making it easier to identify outlier data so that the original problem can be solved without robust weights but also without the outliers.

A robust weighting scheme accounting for the DTI/DKI weights in Eq (8), based on the Geman-McClure

M-estimator, with  $K$  iterations, is [25, 21]:

$$w_n^1 = \left( \exp f_{\theta}(\mathbf{q}_n) \right)^2 \leftarrow \hat{\theta}_{\text{OLS}} \quad (9a)$$

$$w_n^k = \left( \frac{\hat{\sigma} / \exp f_{\theta}(\mathbf{q}_n)}{(\hat{\sigma} / \exp f_{\theta}(\mathbf{q}_n))^2 + u_n^2} \right)^2 \leftarrow \left( \frac{\hat{\theta}_*}{u_n} \right)^{k-1} \quad (9b)$$

$$w_n^{K-1} = \begin{cases} 1 & \text{if } y_n \notin \mathcal{O} \\ 0 & \text{if } y_n \in \mathcal{O} \end{cases} \quad (9c)$$

$$w_n^K = \begin{cases} \left( \exp f_{\theta}(\mathbf{q}_n) \right)^2 \leftarrow \hat{\theta}_*^{K-1} & \text{if } y_n \notin \mathcal{O} \\ 0 & \text{if } y_n \in \mathcal{O} \end{cases} \quad (9d)$$

where  $\hat{\theta}_*^k$  corresponds to the estimated coefficients for iteration  $k$  (we define  $u_n^k$  and  $\hat{\sigma}^k$  below). This scheme requires at least 4 iterations (in which case, a single robustly weighted fit will have been performed for  $k = 2$ ). As with Eq (8), the symbol  $\leftarrow$  in Eq (9) is used to mean that the quantities on the right are used to evaluate the expression for the weights on the left, e.g.,  $w_n^2$  would be calculated using  $\hat{\theta}_*$ ,  $u_n$ , and  $\hat{\sigma}$  from the first iteration. The residuals of the WLS problem are defined as the difference between the log observed signal and log predicted signal:

$$u_n^k := y_n - f_{\theta}(\mathbf{q}_n) \leftarrow \hat{\theta}_*^k \quad (10)$$

The noise level  $\hat{\sigma}$  is estimated at the  $k$ th iteration using a robust estimator designed for the WLS problem [25]:

$$\hat{\sigma}^k = \frac{1.4826}{N - m} N \times \text{MED} [|z_n - \text{MED}[z_n]|] \quad (11)$$

where  $z_n \equiv \exp(f_{\theta}(\mathbf{q}_n))u_n \leftarrow \hat{\theta}_*^k$ , MED is the median operator and  $m$  is the number of regressors, i.e.  $m = 7$  for DTI and  $m = 22$  for DKI.

Set  $\mathcal{O}$  contains log-signals defined as outliers by a 3-sigma rule, applied after the last robustly-weighted fit:

$$y_n \in \mathcal{O} \text{ if } |\exp(y_n) - \exp(f_{\theta}(\mathbf{q}_n))| > 3\hat{\sigma}^{K-2} \quad (12)$$

such that the estimated coefficients  $\hat{\theta}_*$  at iteration  $K - 2$  are used to evaluate  $f_{\theta}(\mathbf{q}_n)$ . If the  $n$ 'th log-signal  $y_n$  is defined as an outlier, then it receives a weight of zero in the last two iterations in Eq (9).

The main insight in this paper is that we can use IRLS with a robust weighting scheme designed specifically for DTI/DKI, i.e., Eq (9), but we are free to choose whether to estimate the coefficients at each iteration using (unconstrained) WLS with Eq (5) or CWLS with Eqs (7). The constraints are independent of the weights, which only enter the cost function Eq (7a) through Eq (4). We will refer to IRLS with weights given by Eq (9) as Robust WLS (RWLS) if Eq (5) is used at each iteration, or as Robust Constrained WLS (RCWLS) if Eq (7) is used at each iteration. The estimated coefficients obtained from the last iteration with weights  $\mathbf{w}^K$  are denoted as  $\hat{\theta}_{\text{RWLS}}$  for RWLS and  $\hat{\theta}_{\text{RCWLS}}$  for RCWLS. For convenience, the unconstrained OLS estimate  $\hat{\theta}_{\text{OLS}}$  is used to define weights for the first iteration for both RWLS and RCWLS. We modified DiPy [26] to be able to solve RWLS and RCWLS. These modifications have been incorporated in DiPy as of v1.10.

## 2.5 Experimental setup and recruitment

Cardiac diffusion-weighted images (cDWI) were acquired on a Connectom 3T research-only MR imaging system (Siemens Healthcare, Erlangen, Germany) with a maximum gradient strength of 300 mT/m and slew

rate of 200 T/m/s. An 18-channel body receive coil was used in combination with a 32-channel spine receive coil. Eleven healthy volunteers (with no known previous cardiac conditions) were recruited for this study: age range  $20.5 \pm 1.9$  years (18 – 24 years), weight range  $64.1 \pm 11.4$  kg (54 – 94 kg), 7 females. The study was approved by the local institutional review board (Cardiff University School of Psychology Research Ethics Committee) and all subjects provided written consent.

A prototype pulse sequence was used that enables diffusion encoding with user-defined second-order motion-compensated (M0M1M2) diffusion gradient waveforms, designed with the NOW toolbox [27, 28, 29, 30] (see Figure A1). The maximum gradient strength used in this study for the M0M1M2-waveform to generate the b-value of  $1350 \text{ s/mm}^2$  was 285.4 mT/m and the maximum physiologically-limited slew-rate was 76.2 T/m/s [15]. The cDWI parameters were: TR = 3 RR-intervals, TE = 61 ms, EPI readout, field-of-view =  $320 \times 120 \text{ mm}^2$  using ZOnally-magnified Oblique Multislice (ZOOM, tilted RF: excitation, tilt angle:  $15^\circ$ , tilted slice thickness: 20 mm) [29, 31], in-plane resolution =  $2.7 \times 2.7 \text{ mm}^2$ , slice thickness = 8 mm, 3 short axis slices (base, mid, and apical), partial Fourier factor = 7/8, no parallel imaging, bandwidth = 2354 Hz/pixel. Each full data set comprised of 5 b-values ( $b = 100, 450, 900, 1200, 1350 \text{ s/mm}^2$ ). For  $b \geq 450 \text{ s/mm}^2$ , 30 directions per shell were acquired with 6 repetitions while  $b = 100 \text{ s/mm}^2$  consisted of 3 directions with 12 repeats. Data were acquired with ECG-gating and under free-breathing (respiratory navigators were not employed for this work). The trigger delay was adjusted for cDWI acquisition in mid-end systole. Saturation bands were placed around the heart. Fat suppression was performed using the SPAIR method [32]. The scan time for all the diffusion weighted images was around 40 minutes depending on subject heart rate. Including cardiac planning, the total scan time was around one hour.

## 2.6 Post-processing

Post-processing was done using in-house tools [20], with rigid image registration utilizing SITK [33] and fitting utilizing DiPy (with our updates) [26]. Image registration was performed by masking a suitable  $b = 100 \text{ s/mm}^2$  image, registering all  $b = 100 \text{ s/mm}^2$  images to this reference image, then using the average of registered  $b = 100 \text{ s/mm}^2$  images to register the entire dataset. Correlation was used as the registration metric, since it outperformed mutual information for high b-value images. The DTI signal representation was then fit to  $b \leq 450 \text{ s/mm}^2$  images using RWLS, and the full image series was predicted. Each original (unmodified) image was then registered to the corresponding predicted image. This method, similar to [34], improved the registration.

After registration, we fit the DTI signal representation to the  $b \leq 450 \text{ s/mm}^2$  data using RWLS. MD, FA, and Helix Angle (HA) (using a cylindrical coordinate system with origin on the LV blood-pool center) were calculated. Segmentation of the LV contours was performed with care taken to exclude voxels exhibiting strong partial-volume effects. For regions strongly affected by artifacts, such as aliasing or susceptibility-induced warping, fitting results do not reliably represent tissue properties. Artifact masks were defined using sectors centered on the LV blood-pool, in order to ignore these parts of the myocardium when calculating voxel statistics. This masking was performed by considering the image series, as well as utilizing MD, FA, HA, root-mean-square-error and coefficient of determination ( $R^2$ ) from the RWLS DTI fit to  $b \leq 450 \text{ s/mm}^2$  data. Across all subjects an average of 25% of voxels were excluded. We have not utilized DKI results to identify artifacts in any way, since this would likely bias comparison between the fitting methods under study.

Having registered the images and segmented the myocardium, we performed further fitting experiments on myocardial voxels only. Defining  $b_{\max}$  as the maximum b-value images that were utilized in a given fit (such that all images with a lower b-value were also included), we performed the following: DTI using WLS and RWLS for  $b_{\max} = 450 \text{ s/mm}^2$ ; DKI with WLS, RWLS, CWLS, and RCWLS, for  $b_{\max}$  values  $900 \text{ s/mm}^2$ ,  $1200 \text{ s/mm}^2$ , and  $1350 \text{ s/mm}^2$ . For RWLS and RCWLS, we used  $K = 10$  iterations. We calculated the following measures in each voxel: mean diffusivity (MD), fractional anisotropy (FA), mean kurtosis (MK), axial kurtosis (AK), radial kurtosis (RK), and radial / axial kurtosis (RK/AK) [2, 4]. We then calculated the average of these measures over non-artifact myocardial voxels.

## 2.7 Statistical methods

In order to make tractable comparisons between different fitting methods, and to isolate the specific effects of constraints and robustness, we applied paired tests to measures obtained from different fitting methods for the same  $b_{\max}$ , and different  $b_{\max}$  for the same fitting method. Specifically, we used the Wilcoxon signed-rank test since non-robust methods often produced results that violate the assumption of normality (which was tested with the Shapiro-Wilk test).

In this work, we potentially face the “multiple comparisons problem”, since we compare many diffusion and kurtosis measures, for many  $b_{\max}$  values and fitting methods. It is not obvious how to adjust significance levels in the context of comparing multiple fitting methods on the same data (particularly given that multiple different measures are calculated from the same set of estimated coefficients, and that higher  $b_{\max}$  fits included all data with lower  $b$ -values). Methods for adjustment of significance levels rely on independence assumptions that do not seem to apply here, especially for a comparison of fitting methods. We therefore take care to draw conclusions that are supported by the overall results.

## 3 Results

### 3.1 Group analysis

Figure 1 shows boxplots of the average DTI/DKI measures for all 11 subjects. All fitting methods are shown for all  $b_{\max}$  values. The black markers drawn along the bottom of each subplot indicate non-normality of the data in each boxplot. Figures 2, 3, and 4 show the results of significance tests between measures from the different fitting methods and different  $b_{\max}$  shown in Figure 1. Figure 2 shows the p-values between different methods within each  $b_{\max}$  in order to demonstrate whether different methods make a significant difference given the same  $b_{\max}$ . Figure 3 shows the p-values between different  $b_{\max}$  within each method, to see whether  $b_{\max}$  made a significant difference in measures given the method. Additionally, Figure 4 shows p-values between DKI methods and DTI methods for MD and FA only, grouped by  $b_{\max}$  of the DKI fits. For  $b = 1350$  s/mm<sup>2</sup>, RCWLS gave the following measures across subjects: MD  $1.68 \pm 0.050 \times 10^{-3}$  mm<sup>2</sup>/s, FA  $0.30 \pm 0.013$ , MK  $0.36 \pm 0.027$ , AK  $0.26 \pm 0.027$ , RK  $0.42 \pm 0.040$ , and RK/AK  $1.65 \pm 0.19$ .

**Constraints:** (i) Adding constraints to a method, i.e., going from WLS to CWLS and from RWLS to RCWLS, made a statistically significant difference on all measures, for all  $b_{\max}$  values, as shown in Figure 2. Figure 1 shows that constraints increased all kurtosis measures MK, AK, RK, and RK/AK, the latter results showing that constraints result in increasing RK more than AK. Increased kurtosis is consistent with the constraints ensuring that fitted parameters correspond to non-negative kurtosis. (ii) For constrained fitting, MK, RK, and AK all decrease with  $b_{\max}$ , with Figure 3 showing these changes are significant for both CWLS and RCWLS. However, Figure 1 indicates that the ratio RK/AK appears to increase with  $b_{\max}$  for both constrained and unconstrained fitting, although this difference is not significant between  $b_{\max} = 1200$  s/mm<sup>2</sup> and  $b_{\max} = 1350$  s/mm<sup>2</sup>.

**Robustness:** (i) Robust fitting by itself, i.e., going from WLS to RWLS, and CWLS to RCWLS, gave large changes in measures for some subjects (in particular, reducing MD and MK) but did not significantly change the mean measure values over subjects. Robust fitting generally reduces the spread of the measures over the group by correcting errors in measures for some subjects, as is visually clear in Figure 1. Non-normality of measures was only found for non-robust methods (with the *single exception*, out of such 36 tests, being AK for RCWLS at  $b_{\max} = 1350$  s/mm<sup>2</sup>). (ii) Robustness *may* have an effect on RK/AK; Figure 2 shows the following p-values comparing CWLS and RCWLS:  $b_{\max} = 900$  s/mm<sup>2</sup> ( $p = 0.032$ ),  $b_{\max} = 1200$  s/mm<sup>2</sup> ( $p = 0.068$ ), and  $b_{\max} = 1350$  s/mm<sup>2</sup> ( $p = 0.024$ ). However, in the context of multiple comparisons, where no effect of robustness was seen on other measures and a 0.05 significance threshold was not met for all  $b_{\max}$ , this result cannot be ascribed significance.

**MD and FA:** for RWLS DTI fits, we obtained MD  $1.53 \pm 0.074 \times 10^{-3}$  mm<sup>2</sup>/s and FA  $0.31 \pm 0.017$ . When considering robust methods, (i) MD increases from DTI ( $b_{\max} \leq 450$  s/mm<sup>2</sup>) to unconstrained DKI (see Figure 4), and from unconstrained DKI to constrained DKI (see Figure 2); FA increases from DTI to unconstrained DKI (see Figure 4) but constrained DKI results in lower FA than DTI fitting (see Figure 4).



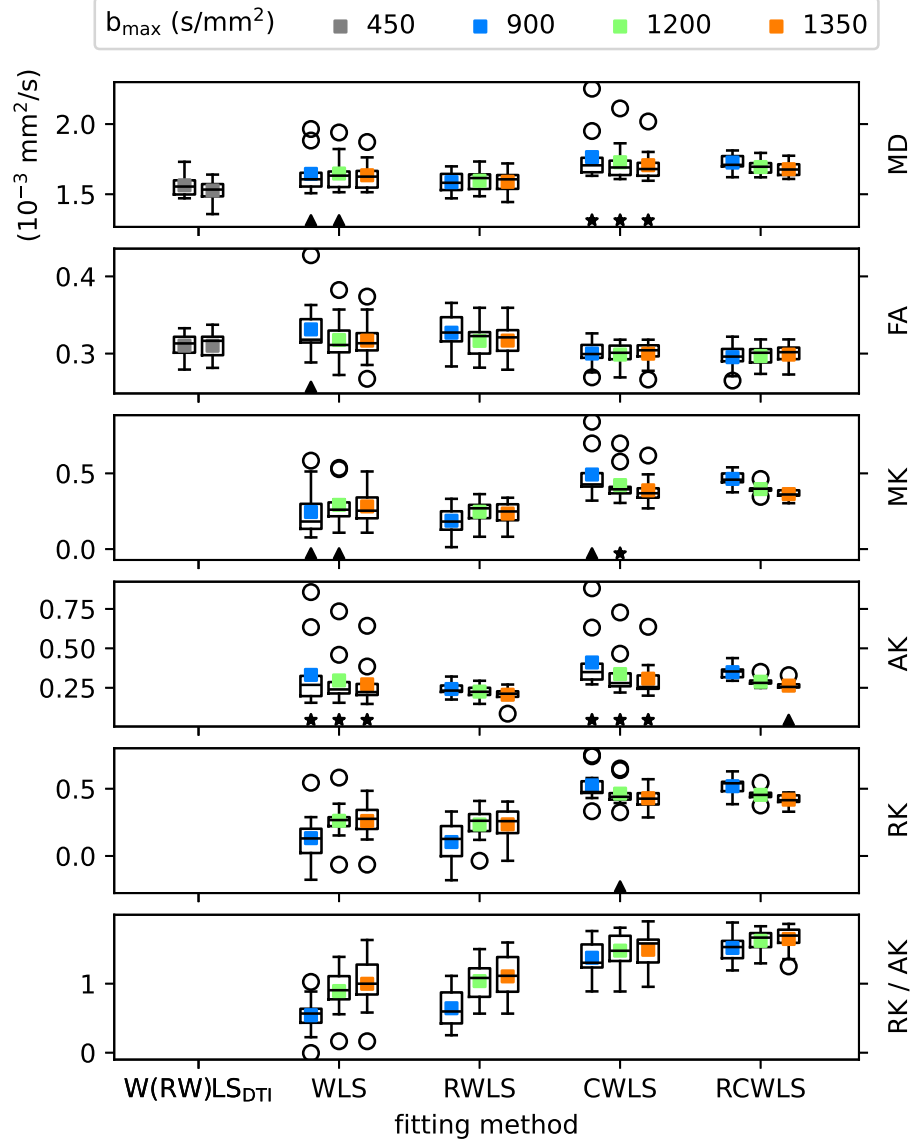


Figure 1: Average of DTI/DKI measures over myocardial voxels for all subjects. Colored points show the average over subjects, colored by  $b_{\max}$ . DTI fits W(RW)LS<sub>DTI</sub> show WLS (left) and RWLS (right) together for convenience. The black triangles (stars) show where the Shapiro-Wilk test p-value  $\leq 0.05$  ( $\leq 0.01$ ), i.e., the hypothesis of normality is rejected.

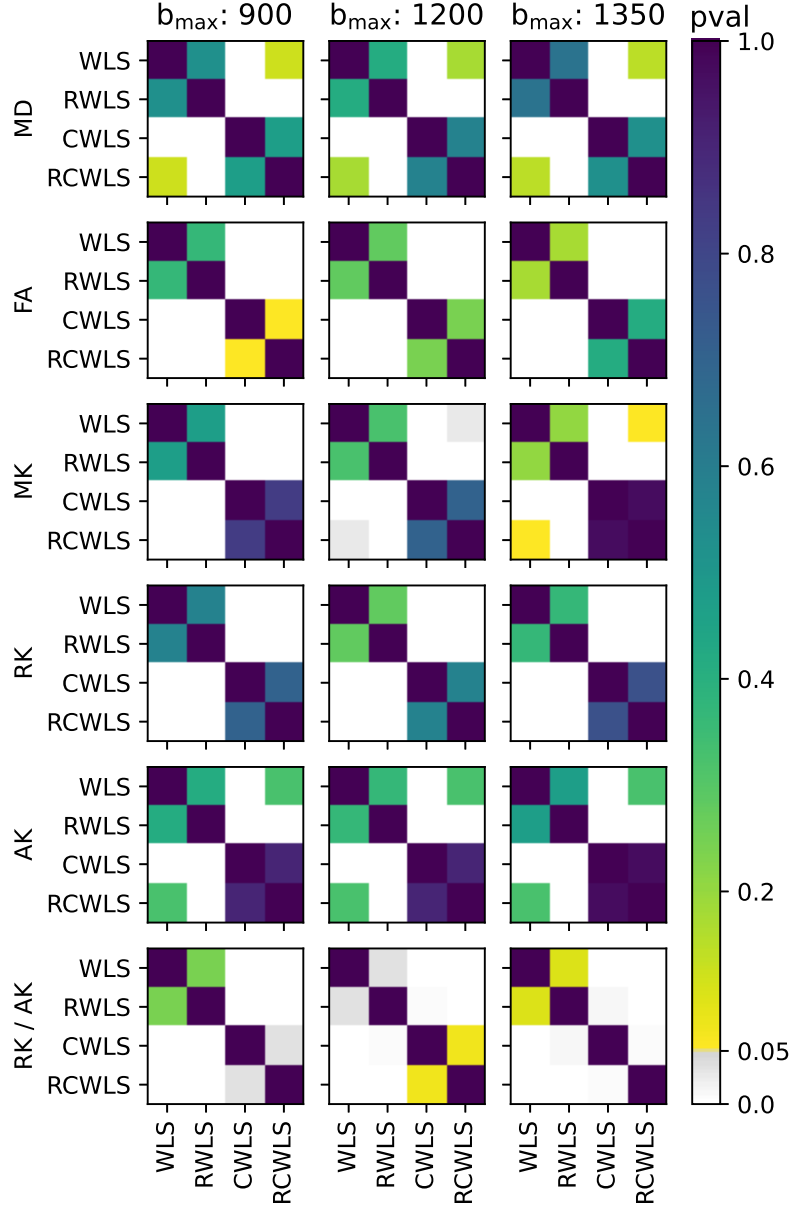


Figure 2: p-values from comparing DKI measures from different fitting methods given the same  $b_{\max}$  (units  $s/mm^2$ ).

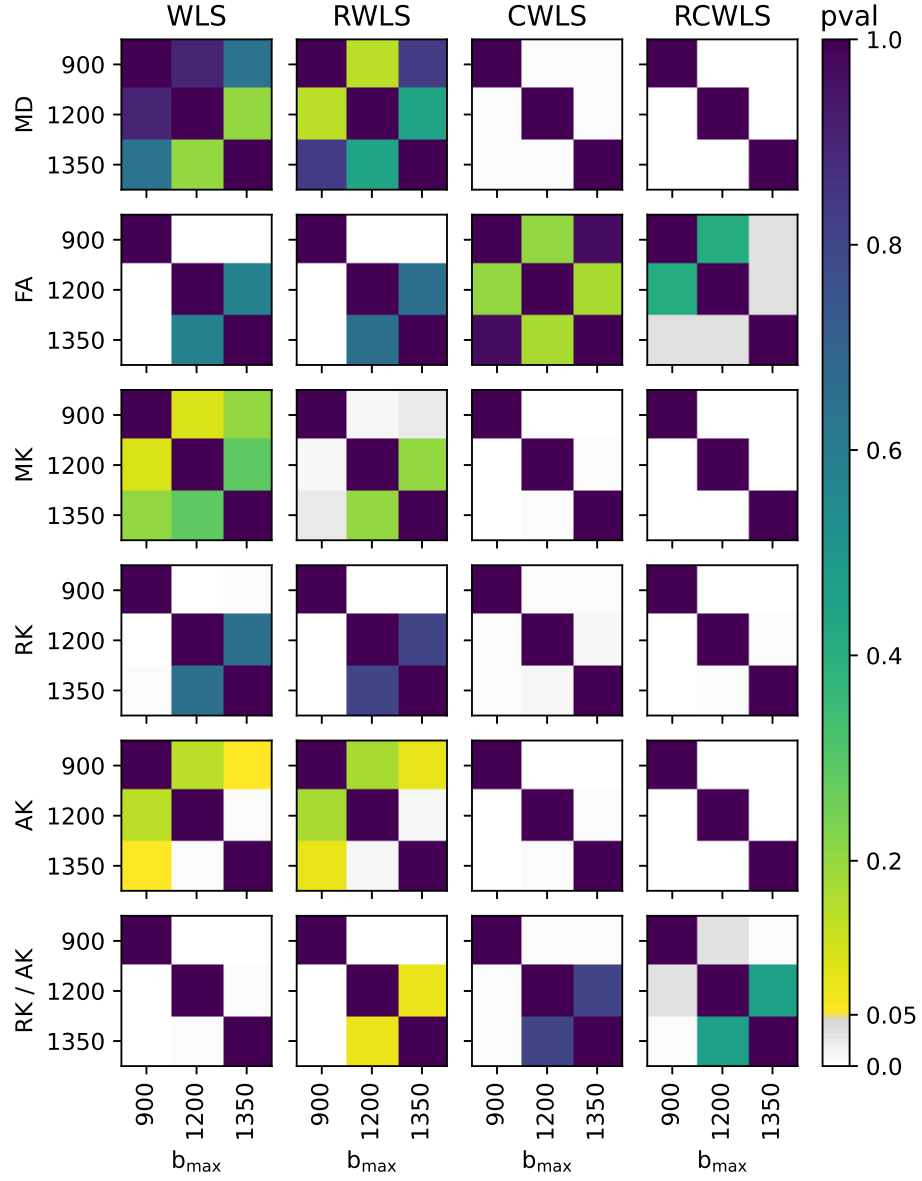


Figure 3: p-values from comparing DKI measures from different  $b_{\max}$  (units  $\text{s}/\text{mm}^2$ ) given the same fitting methods.

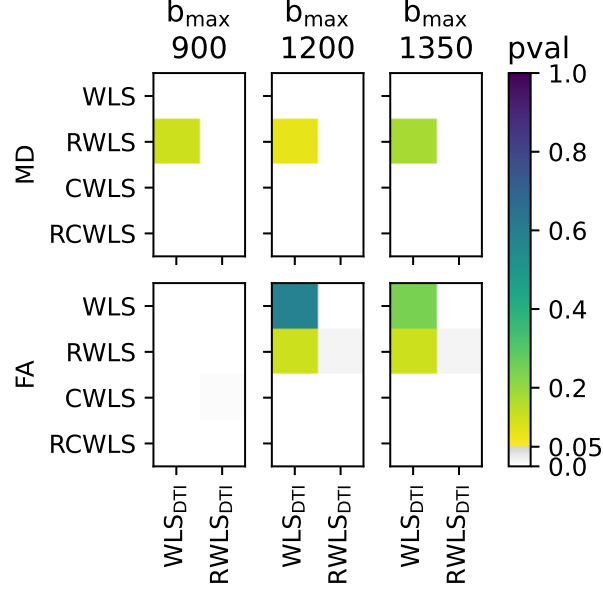


Figure 4: p-values from comparing MD and FA from DKI fitting methods for each  $b_{\max}$  used for DKI fitting (units  $\text{s/mm}^2$ ) against DTI fitting methods ( $b_{\max} = 450 \text{ s/mm}^2$ ).

(ii) For constrained methods, as  $b_{\max}$  increases MD decreases slightly but there is little change in FA (see Figure 3).

### 3.2 Example maps

Figures 5, 6, 7, and 8 show example measure maps for 4 different subjects. The colormap ranges were chosen based on the boxplots in Figure 1, in particular to emphasize whether values are above zero for MK or above one for RK/AK.

Figure 5 shows basal-slice maps for  $b_{\max} = 1350 \text{ s/mm}^2$  for the subject with the second highest MD and MK for non-robust fitting WLS/CWLS in Figure 1. Robust fitting gave large changes over the whole myocardium, bringing measures into better agreement with other subjects, such that for robust fitting methods this subject is no longer an outlier in the boxplots of Figure 1. MD and MK are reduced, and FA is increased, for robust methods vs non-robust methods. Constraints increase MK and RK/AK, which is most noticeable between RWLS and RCWLS. In particular for RCWLS,  $\text{RK/AK} > 1$  in nearly every voxel and is relatively homogeneous. This example shows that kurtosis can appear to be positive for non-constrained methods, but these kurtosis measures can still be corrupted. Going from WLS to RWLS results in a large reduction of MK, indicating that its consistently positive value for WLS was likely due to corruptions in the data (which, in this case, also caused an inflation of MD). By adding constraints to robust fitting, i.e. going from RWLS to RCWLS, consistently positive MK is recovered, but with much lower values than from the WLS fit.

Figure 6 shows mid-slice maps for  $b_{\max} = 1350 \text{ s/mm}^2$  where robust fitting had large effects in a localized region (around 2 to 6 o'clock) with low MD, high FA, and negative MK for WLS. Here, both robust fitting and constraints independently increase MD, reduce FA, and increase MK, such that RCWLS gives the most spatially uniform values. The regions where RK/AK is around 1 for RCWLS coincide with relatively lower FA.

Figure 7 shows MK and RK/AK basal-slice maps for different  $b_{\max}$ . Constraints give positive MK and make  $\text{RK/AK} > 1$  overall. Reading across the rows, constraints increase MK and RK / AK given the same  $b_{\max}$ . Differences between unconstrained and constrained fitting decrease with  $b_{\max}$ . Reading down the

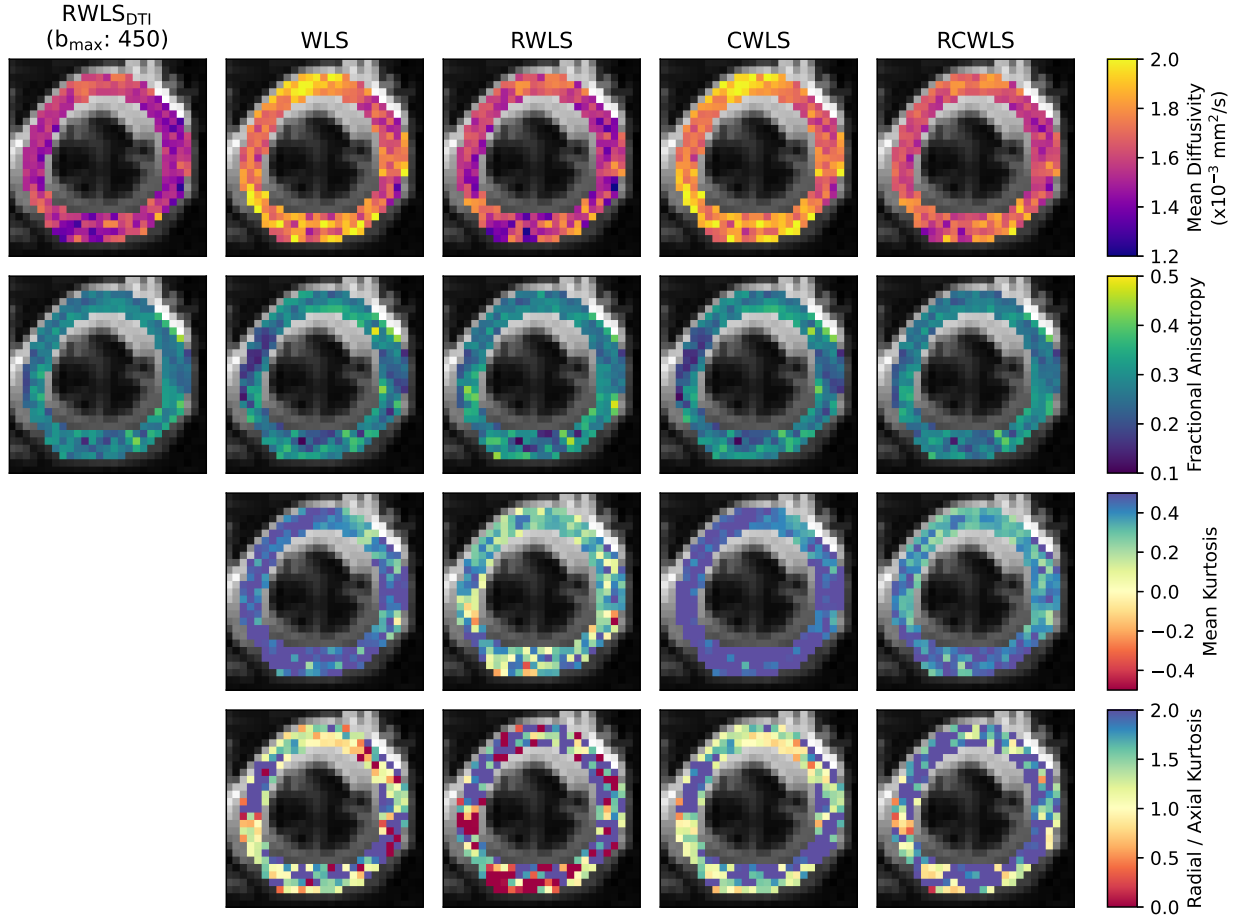


Figure 5: Maps for the mid slice of a single subject, for  $b_{\max} = 1350 \text{ s/mm}^2$ . The first column shows MD and FA from DTI RWLS fitting ( $b_{\max} = 450 \text{ s/mm}^2$ ) for reference. All other columns show measures from DKI fits. In this example, robust fitting decreases MK, while constrained fitting increases MK. Using robust and constrained fitting (RCWLS) gives the most plausible results, with mostly RK/AK  $> 1$  values.

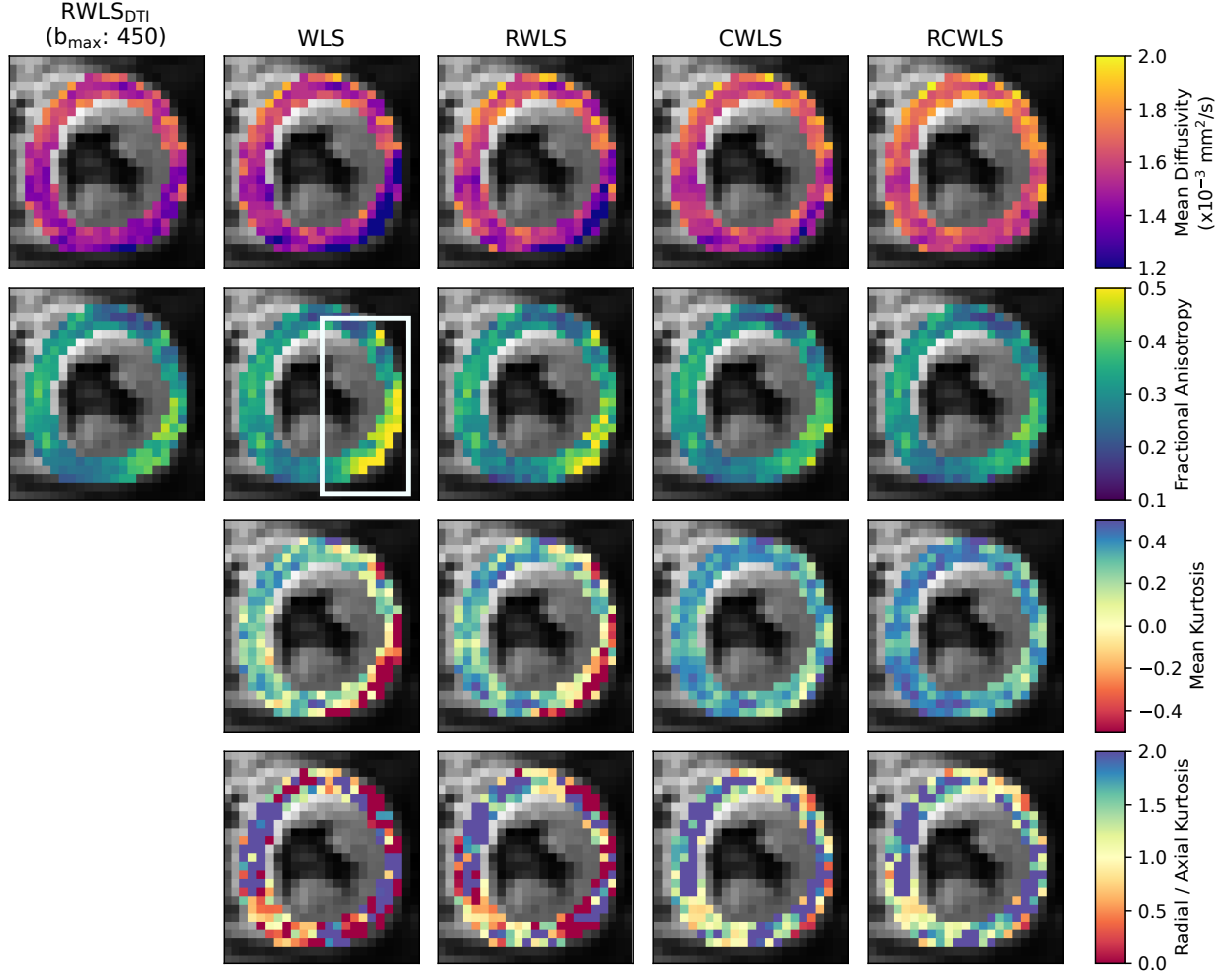


Figure 6: Measure maps for the mid slice of a single subject, for  $b_{\max} = 1350 \text{ s/mm}^2$ . The first column shows MD and FA from DTI RWLS fitting ( $b_{\max} = 450 \text{ s/mm}^2$ ) for reference. All other columns show measures from DKI fits. Around 2 to 6 o'clock for the DKI fits is a region of implausibly low MD, high FA, and negative MK, for unconstrained fits. Constraints correct this region towards plausible values, with robust fitting visibly improving the region further in the same direction.

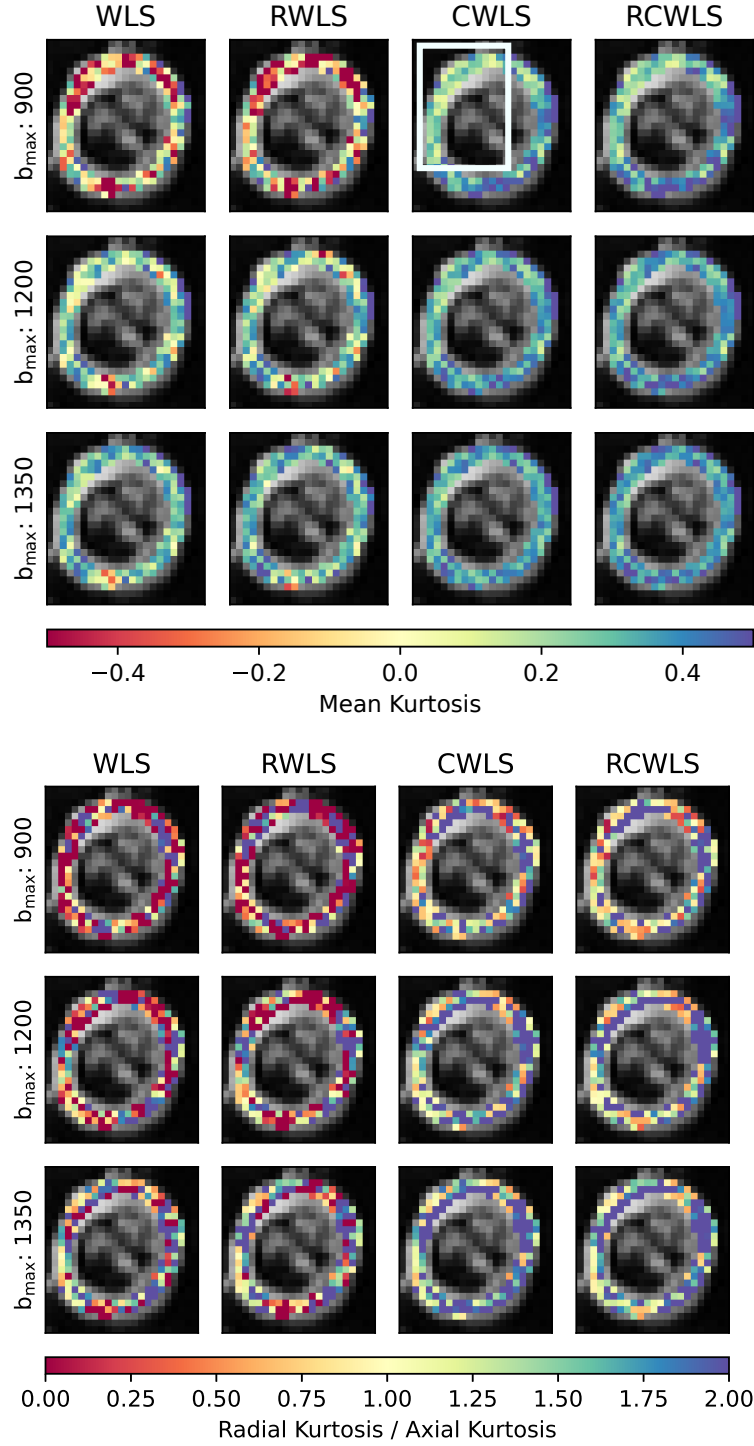


Figure 7: Mean Kurtosis (MK) and Radial Kurtosis / Axial Kurtosis (RK/AK) maps for the basal slice of a single subject, for all  $b_{\max}$  values (units  $\text{s/mm}^2$ ) used for DKI fitting methods. Constraints increase MK given the same  $b_{\max}$ . For constrained fitting (CWLS and RCWLS), MK decreases with  $b_{\max}$  in most voxels but increases between 8 and 12 o'clock. RCWLS for  $b_{\max} = 1350 \text{ s/mm}^2$  has the most spatially homogeneous MK.

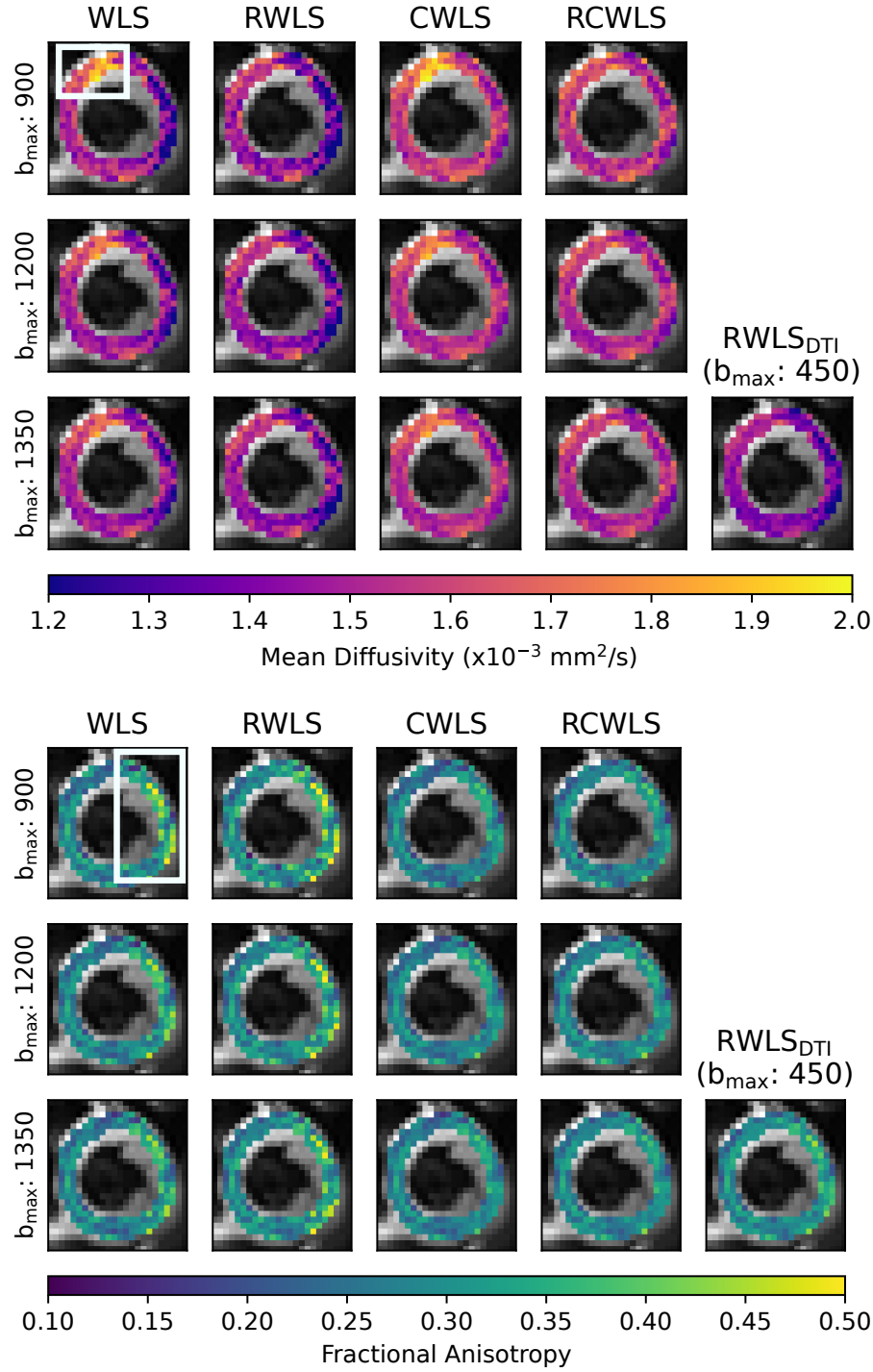


Figure 8: Mean Diffusivity (MD) and Fractional Anisotropy (FA) maps for the basal slice of a single subject, for all  $b_{\max}$  values (units  $\text{s}/\text{mm}^2$ ) used for DKI fitting methods. Also shown are MD and FA from RWLS DTI fitting ( $b_{\max} = 450 \text{ s}/\text{mm}^2$ ) for reference. Robust fitting corrects elevated MD in the top-left (10 - 12 o'clock), but constraints are needed to correct the reduced MD and elevated FA on the right (12 - 5 o'clock).



RCWLS column, MK decreases with  $b_{\max}$  for most voxels but increases between 8 and 12 o'clock. These opposing effects result in RCWLS for  $b_{\max} = 1350$  s/mm<sup>2</sup> having the most homogeneous MK.

Figure 8 shows MD and FA maps in a basal slice for different  $b_{\max}$  and for DTI fitting ( $b_{\max} = 450$  s/mm<sup>2</sup>). Robust fitting corrects elevated MD in the top-left (10 - 12 o'clock), but constraints are needed to correct the reduced MD and elevated FA on the right (12 - 5 o'clock). The effects on MD and FA are consistent with the overall group trends. RCWLS gives the most homogeneous MD and FA, with the least variation with  $b_{\max}$ .

## 4 Discussion

Our results show that both robust fitting and convexity constraints affect DTI/DKI measures in important ways. Figure 5 helps to show that the RCWLS results are not just the robust result with the negative kurtosis turned to zero - the kurtosis becomes convincingly positive, and all measures change when adding constraints. In this case, robust fitting reduced inflated MK caused by corruptions, while adding constraints increased MK. Figure 6 demonstrates a region where only robust fitting and constraints together appear to fully resolve a region of corrupted measures. In theory, RCWLS provides advantages greater than the sum of its parts: convexity constraints should make outlier identification easier, which should improve the weights in the final WLS fit.

**Constraints:** Imposing convexity constraints resulted in statistically significant differences in all measures for all  $b_{\max}$  values. The precise changes, notably increased MK and RK/AK, are important. Only constrained fitting reliably shows  $RK > AK$  (for RCWLS  $b_{\max} = 1350$  s/mm<sup>2</sup>, RK/AK mean and median are 1.65 and 1.70 respectively), despite this being expected in myocardium since there are fewer restrictions to diffusion parallel to the long axes of cardiomyocytes [35, 3]. It is worth emphasizing that adding constraints does not just turn negative kurtosis into zero kurtosis in certain voxels (something that could be trivially obtained in post-processing the fitting results), but fundamentally alters the optimum coefficient vector (Figures 5 - 6).

**Robustness:** Robust fitting can have large overall and regional benefits on individual subjects, which is invaluable for potential clinical applications. Variation between subjects within a group ought to represent physiological variation (and non-gross measurement error) rather than the effects of image corruptions. In the context of a study comparing different groups (e.g. healthy volunteers vs disease), the reduction in the spread of measures from robust fitting can increase statistical power and lower the probability of a type 2 errors (false negatives) [21].

**MD and FA:** While it is known that MD and FA obtained from fitting the DKI signal representation can be different compared to those obtained from the DTI signal representation [36], constraints further modify these values: MD increases from DTI to DKI and from unconstrained to constrained fitting, but while FA increases from DTI to DKI, it decreases from unconstrained to constrained fitting, resulting in RCWLS giving the lowest FA values. This shows that the FA changes from DTI to DKI were largely in the context of parameter fits that violated constraints. To the best of our knowledge, other works noting differences in MD and FA between DTI and DKI have not utilized constraints or robust fitting.

**$b_{\max}$  effect:** Varying  $b_{\max}$  allowed for investigating the importance of constraints and robustness for different experimental designs: higher  $b_{\max}$  data is more informative about kurtosis but has lower SNR. While previous work has noted the dependence of the performance of (simplified models of) cardiac DKI on  $b_{\max}$  [3], it is noteworthy that MK and RK appear to increase with  $b_{\max}$  for unconstrained fitting, but decrease with  $b_{\max}$  for constrained fitting. This only serves to emphasize the importance of using fitting constraints. Due to these opposite trends, the differences between unconstrained and constrained fitting get smaller as  $b_{\max}$  increases, suggesting fewer and milder constraint violations as data become more informative about kurtosis.

## Limitations

Although the differences between non-robust and robust methods were generally insignificant at the group level, the errors from non-robust methods would make the detection of any actually existing differences quite challenging, which is perhaps ironic. It is also true that summary statistics such as mean measures over all voxels can be insensitive to the improvement in measure maps when using robust fitting, and yet post-hoc statistical analysis of regions of interest (where measures were most changed by robust methods) would also seem problematic. The only measure that indicated any *potential* effect (at the group level) from robust fitting alone was RK/AK, particularly for  $b_{\max} = 1350 \text{ s/mm}^2$  ( $p = 0.024$ ), but we cannot rule out a false positive here due to multiple comparisons of many measures. The ratio RK/AK might be more sensitive to changes from robust fitting, but more subjects would be required to have sufficient statistical power to determine this.

Higher  $b_{\max}$  fits included all lower  $b$ -values, and so there is more data available for these fits. Our study sought to determine how the results changed as more information about kurtosis was available, specifically in the context of understanding the effects of robust fitting and constrained fitting, so the conflation from having both higher  $b$ -value data (which has lower SNR) and more data overall was not of particular concern. To perform an analysis about the suitability of different data designs, we believe many more subjects would be required, as well as a more nuanced analysis of trends with  $b_{\max}$  (rather than just paired tests). This work could be restricted to robust and constrained fitting methods.

A limitation of signal representations, such as DTI and DKI, is a lack of clear insight into the specific quantitative links to the underlying biology: differences in kurtosis measures between groups can be observed and reasoned about, but the absolute values are not ascribed any particular meaning [3]. Nonetheless, we can speculate that kurtosis measures (derived from appropriate fitting techniques) may be a way to increase the biomarker space and gain insight into disease; further work is required.

## 5 Conclusion

In this work, we have developed robust constrained weighted least squares (RCWLS), the first robust estimation technique for DKI that incorporates necessary constraints on the signal behavior. Using in vivo human cardiac DKI data from healthy volunteers collected with a Connectom scanner, we determined that RCWLS is the most suitable fitting technique compared with others that lack either robustness or constraints. For  $b = 1350 \text{ s/mm}^2$ , RCWLS gave the following measures across subjects: MD  $1.68 \pm 0.050 \times 10^{-3} \text{ mm}^2/\text{s}$ , FA  $0.30 \pm 0.013$ , MK  $0.36 \pm 0.027$ , AK  $0.26 \pm 0.027$ , RK  $0.42 \pm 0.040$ , and RK/AK  $1.65 \pm 0.19$ . Constraints, but not robustness, had a significant effect on all diffusion and kurtosis measures. However, robust fitting corrected large errors for some subjects and generally improved diffusion and kurtosis maps. Only RCWLS convincingly showed radial kurtosis to be larger than axial kurtosis for all subjects, something that is expected in myocardium due to increased restrictions to diffusion in the plane perpendicular to the primary myocyte direction. RCWLS also showed the best correction of corrupted regions in diffusion parameter maps for individual subjects. Future work on in vivo cardiac DKI should utilize fitting techniques that are both robust and constrained, such as RCWLS.

## A Gradient waveforms

Figure A1 shows the gradient waveforms used in this work. The timing of the diffusion encoding waveform is 17, 8, and 17 ms for the pre-, pause- and post-duration, respectively.

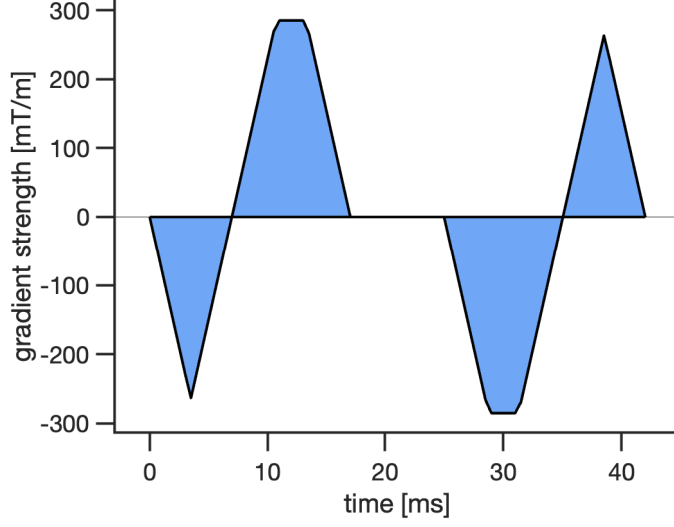


Figure A1: Numerically optimized second-order motion compensated waveform for  $b_{\max} = 1350 \text{ s/mm}^2$ ,  $G_{\max} = 285.4 \text{ mT/m}$  and maximum slew rate  $76.2 \text{ T/m/s}$ .

## B Constraint matrices

Many basis elements can be excluded a priori because they cannot contribute, leaving:

$$\mathbf{e} = \left( s_1, s_2, s_3, q_1 s_1, q_1 s_2, q_1 s_3, \right. \\ \left. q_2 s_1, q_2 s_2, q_2 s_3, q_3 s_1, q_3 s_2, q_3 s_3 \right)^T \quad (\text{B1})$$

We can parameterize the matrix  $H(\boldsymbol{\theta})$  as:

$$2 \begin{bmatrix} \theta_1 & \theta_2 & \theta_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \theta_2 & \theta_3 & \theta_5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \theta_4 & \theta_5 & \theta_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \theta_7 & \theta_{10} & \theta_{11} & \theta_{10} & \theta_{16} & \theta_{19} & \theta_{11} & \theta_{19} & \theta_{17} \\ 0 & 0 & 0 & \theta_{10} & \theta_{16} & \theta_{19} & \theta_{16} & \theta_{12} & \theta_{20} & \theta_{19} & \theta_{20} & \theta_{21} \\ 0 & 0 & 0 & \theta_{11} & \theta_{19} & \theta_{17} & \theta_{19} & \theta_{20} & \theta_{21} & \theta_{17} & \theta_{21} & \theta_{14} \\ 0 & 0 & 0 & \theta_{10} & \theta_{16} & \theta_{19} & \theta_{16} & \theta_{12} & \theta_{20} & \theta_{19} & \theta_{20} & \theta_{21} \\ 0 & 0 & 0 & \theta_{16} & \theta_{12} & \theta_{20} & \theta_{12} & \theta_8 & \theta_{13} & \theta_{20} & \theta_{13} & \theta_{18} \\ 0 & 0 & 0 & \theta_{19} & \theta_{20} & \theta_{21} & \theta_{20} & \theta_{13} & \theta_{18} & \theta_{21} & \theta_{18} & \theta_{15} \\ 0 & 0 & 0 & \theta_{11} & \theta_{19} & \theta_{17} & \theta_{19} & \theta_{20} & \theta_{21} & \theta_{17} & \theta_{21} & \theta_{14} \\ 0 & 0 & 0 & \theta_{19} & \theta_{20} & \theta_{21} & \theta_{20} & \theta_{13} & \theta_{18} & \theta_{21} & \theta_{18} & \theta_{15} \\ 0 & 0 & 0 & \theta_{17} & \theta_{21} & \theta_{14} & \theta_{21} & \theta_{18} & \theta_{15} & \theta_{14} & \theta_{15} & \theta_9 \end{bmatrix} \quad (\text{B2})$$

The ‘slack matrix’  $L(\boldsymbol{\alpha})$  is determined by Eq (7f). We can write  $L(\boldsymbol{\alpha}) = \sum_{u=1} \alpha_u L_u$ , such that  $\mathbf{e}^T \cdot L_u \cdot \mathbf{e} = 0$ . There are 18 such matrices, so we parameterize  $L(\boldsymbol{\alpha})$  as:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & \beta_{10} & \beta_{11} & 0 & \beta_{13} & \beta_{14} & 0 & \beta_{16} & \beta_{17} \\ 0 & 0 & 0 & \alpha_{10} & 0 & \beta_{12} & \alpha_{13} & 0 & \beta_{15} & \alpha_{16} & 0 & \beta_{18} \\ 0 & 0 & 0 & \alpha_{11} & \alpha_{12} & 0 & \alpha_{14} & \alpha_{15} & 0 & \alpha_{17} & \alpha_{18} & 0 \\ 0 & \alpha_{10} & \alpha_{11} & 0 & 0 & 0 & 0 & \beta_1 & \beta_2 & 0 & \beta_4 & \beta_5 \\ \beta_{10} & 0 & \alpha_{12} & 0 & 0 & 0 & \alpha_1 & 0 & \beta_3 & \alpha_4 & 0 & \beta_6 \\ \beta_{11} & \beta_{12} & 0 & 0 & 0 & 0 & \alpha_2 & \alpha_3 & 0 & \alpha_5 & \alpha_6 & 0 \\ 0 & \alpha_{13} & \alpha_{14} & 0 & \alpha_1 & \alpha_2 & 0 & 0 & 0 & 0 & \beta_7 & \beta_8 \\ \beta_{13} & 0 & \alpha_{15} & \beta_1 & 0 & \alpha_3 & 0 & 0 & 0 & \alpha_7 & 0 & \beta_9 \\ \beta_{14} & \beta_{15} & 0 & \beta_2 & \beta_3 & 0 & 0 & 0 & 0 & \alpha_8 & \alpha_9 & 0 \\ 0 & \alpha_{16} & \alpha_{17} & 0 & \alpha_4 & \alpha_5 & 0 & \alpha_7 & \alpha_8 & 0 & 0 & 0 \\ \beta_{16} & 0 & \alpha_{18} & \beta_4 & 0 & \alpha_6 & \beta_7 & 0 & \alpha_9 & 0 & 0 & 0 \\ \beta_{17} & \beta_{18} & 0 & \beta_5 & \beta_6 & 0 & \beta_8 & \beta_9 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (\text{B3})$$

where  $\beta_{ij} = -\alpha_{ij}$  (where  $\beta_{ij}$  are only used here to assist with display, they are not additional parameters).

The reason for writing the sum of squares representation as matrix  $H(\boldsymbol{\theta}) + L(\boldsymbol{\alpha})$  in Eq (7b), and therefore introducing parameters  $\boldsymbol{\alpha}$ , can be understood by considering (7e): there are (infinitely) many choices for  $H(\boldsymbol{\theta})$  given basis Eq (B1), but we must implement a specific matrix  $H(\boldsymbol{\theta})$  (e.g. Eq (B2)). The flexibility is then explicitly parameterized via  $L(\boldsymbol{\alpha})$ .

## C Reduced constrained problem

As explained in the appendix of [19], the size of the semidefinite programming problem can be reduced (giving significant computational savings) by replacing Eq (7a) with

$$\arg \min_{\boldsymbol{\theta}, \boldsymbol{\alpha}} \left\| \mathbf{V}'^T \cdot \boldsymbol{\theta} - \mathbf{V}'^T \cdot \hat{\boldsymbol{\theta}}_{\text{WLS}} \right\|^2 \quad (\text{C1})$$

where  $\hat{\boldsymbol{\theta}}_{\text{WLS}}$  is the unconstrained WLS solution given by Eq (5) and  $\mathbf{V}$  is given by the Cholesky factorization  $(\mathbf{X}'^T \cdot \mathbf{X}') = \mathbf{V} \cdot \mathbf{V}^T$ . For RCWLS the problem is reduced at each iteration (the *unconstrained* fit result  $\hat{\boldsymbol{\theta}}_{\text{WLS}}$  depends on the previous *constrained* fit). The semi-definite program Eq (7) only needs solving if the WLS estimate violates the constraints.

## Acknowledgments

S. Coveney would like to thank Serge Koudoro, Ariel Rokem, and Rafael Neto Henriques for help with implementing our work in DiPy, and Tom Dela Haije for advice on deriving the constraint matrices.

## Financial disclosure

None reported.

## Conflict of interest

The authors declare no potential conflict of interests.

## References

- [1] D. Alexander, G. Barker, and S. Arridge, “Detection and modeling of non-gaussian apparent diffusion coefficient profiles in human brain data,” *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 48, no. 2, pp. 331–340, 2002.

- [2] J. H. Jensen, J. A. Helpert, A. Ramani, H. Lu, and K. Kaczynski, "Diffusional kurtosis imaging: the quantification of non-gaussian water diffusion by means of magnetic resonance imaging," *Magnetic Resonance in Medicine*, vol. 53, no. 6, pp. 1432–1440, Jun. 2005.
- [3] D. McClymont, I. Teh, E. Carruth, J. Omens, A. McCulloch, H. J. Whittington, P. Kohl, V. Grau, and J. E. Schneider, "Evaluation of non-Gaussian diffusion in cardiac MRI," *Magnetic resonance in medicine*, vol. 78, no. 3, pp. 1174–1186, 2017.
- [4] J. H. Jensen and J. A. Helpert, "MRI quantification of non-Gaussian water diffusion by kurtosis analysis," *NMR in biomedicine*, vol. 23, no. 7, pp. 698–710, Aug. 2010.
- [5] A. Tabesh, J. H. Jensen, B. A. Ardekani, and J. A. Helpert, "Estimation of tensors and tensor-derived measures in diffusional kurtosis imaging," *Magnetic resonance in medicine*, vol. 65, no. 3, pp. 823–836, 2011.
- [6] P. J. Basser and C. Pierpaoli, "Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor mri," *Journal of magnetic resonance*, vol. 213, no. 2, pp. 560–570, 2011.
- [7] G. R. Glenn, J. A. Helpert, A. Tabesh, and J. H. Jensen, "Quantitative assessment of diffusional kurtosis anisotropy," *NMR in Biomedicine*, vol. 28, no. 4, pp. 448–459, 2015.
- [8] B. Hansen, N. Shemesh, and S. N. Jespersen, "Fast imaging of mean, axial and radial diffusion kurtosis," *Neuroimage*, vol. 142, pp. 381–393, 2016.
- [9] B. Hansen, T. E. Lund, R. Sangill, and S. N. Jespersen, "Experimentally and computationally fast method for estimation of a mean kurtosis," *Magnetic resonance in medicine*, vol. 69, no. 6, pp. 1754–1760, 2013.
- [10] S. N. Jespersen, "White matter biomarkers from diffusion mri," *Journal of Magnetic Resonance*, vol. 291, pp. 127–140, 2018.
- [11] R. N. Henriques, M. M. Correia, R. G. Nunes, and H. A. Ferreira, "Exploring the 3d geometry of the diffusion kurtosis tensor—impact on the development of robust tractography procedures and novel biomarkers," *Neuroimage*, vol. 111, pp. 85–99, 2015.
- [12] R. N. Henriques, S. N. Jespersen, D. K. Jones, and J. Veraart, "Toward more robust and reproducible diffusion kurtosis imaging," *Magnetic resonance in medicine*, vol. 86, no. 3, pp. 1600–1613, 2021.
- [13] I. Teh, D. Shelley, J. H. Boyle, F. Zhou, A.-M. Poenar, N. Sharrack, R. J. Foster, N. Y. Yuldasheva, G. J. Parker, E. Dall’Armellina *et al.*, "Cardiac q-space trajectory imaging by motion-compensated tensor-valued diffusion encoding in human heart in vivo," *Magnetic Resonance in Medicine*, vol. 90, no. 1, pp. 150–165, 2023.
- [14] C. A. Hanson, A. Kamath, M. Gottbrecht, S. Ibrahim, and M. Salerno, "T2 relaxation times at cardiac mri in healthy adults: a systematic review and meta-analysis," *Radiology*, vol. 297, no. 2, pp. 344–351, 2020.
- [15] M. Afzali, L. Mueller, S. Coveney, F. Fasano, C. J. Evans, M. Engel, F. Szczepankiewicz, I. Teh, E. Dall’Armellina, D. K. Jones, and J. E. Schneider, "In vivo diffusion MRI of the human heart using a 300 mT/m gradient system," *Magnetic Resonance in Medicine*, vol. 92, no. 3, pp. 1022–1034, 2024.
- [16] M. Afzali, L. Mueller, S. Coveney, S. Jones, F. Fasano, J. Evans, I. Teh, E. Dall’Armellina, F. Szczepankiewicz, D. K. Jones, and J. E. Schneider, "Quantification of non-Gaussian diffusion in the human heart in vivo," *Proceedings of the Annual Meeting of ISMRM*, 2024.

- [17] M. Afzali, S. Coveney, L. Mueller, S. Jones, F. Fasano, C. J. Evans, I. Teh, E. Dall’Armellina, F. Szczepankiewicz, D. K. Jones, and J. E. Schneider, “Cardiac diffusion kurtosis imaging in the human heart in vivo using 300 mT/m gradients,” *Magnetic Resonance in Medicine*, Jul. 2025, publisher: Wiley. [Online]. Available: <https://onlinelibrary.wiley.com/doi/10.1002/mrm.30626>
- [18] C. M. Tax, W. M. Otte, M. A. Viergever, R. M. Dijkhuizen, and A. Leemans, “REKINDLE: Robust extraction of kurtosis INDices with linear estimation: Rekindle,” *Magnetic Resonance in Medicine*, vol. 73, no. 2, pp. 794–808, Feb. 2015. [Online]. Available: <https://onlinelibrary.wiley.com/doi/10.1002/mrm.25165>
- [19] T. Dela Haije, E. Özarslan, and A. Feragen, “Enforcing necessary non-negativity constraints for common diffusion MRI models using sum of squares programming,” *NeuroImage*, vol. 209, p. 116405, Apr. 2020. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1053811919309966>
- [20] S. Coveney, C. Kelly, I. Teh, M. Afzali, L. Mueller, D. Arka, F. Szczepankiewicz, D. K. Jones, E. Dall’Armellina, and J. E. Schneider, “Semi-Automated Rejection of Corrupted Images in Cardiac Diffusion Tensor Imaging,” *Proceedings of the Annual Meeting of ISMRM*, 2023.
- [21] S. Coveney, M. Afzali, L. Mueller, I. Teh, A. Das, E. Dall’Armellina, F. Szczepankiewicz, D. K. Jones, and J. E. Schneider, “Outlier detection in cardiac diffusion tensor imaging: Shot rejection or robust fitting?” *Medical Image Analysis*, vol. 101, p. 103386, Apr. 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1361841524003116>
- [22] J. Veraart, W. Van Hecke, and J. Sijbers, “Constrained maximum likelihood estimation of the diffusion kurtosis tensor using a Rician noise model,” *Magnetic Resonance in Medicine*, vol. 66, no. 3, pp. 678–686, Sep. 2011.
- [23] A. Magnani, S. Lall, and S. Boyd, “Tractable fitting with convex polynomials via sum-of-squares,” in *Proceedings of the 44th IEEE Conference on Decision and Control*, Dec. 2005, pp. 1672–1677, iSSN: 0191-2216. [Online]. Available: <https://ieeexplore.ieee.org/document/1582399>
- [24] R. Salvador, A. Peña, D. K. Menon, T. A. Carpenter, J. D. Pickard, and E. T. Bullmore, “Formal characterization and extension of the linearized diffusion tensor model,” *Human Brain Mapping*, vol. 24, no. 2, pp. 144–155, Feb. 2005.
- [25] Q. Collier, J. Veraart, B. Jeurissen, A. J. den Dekker, and J. Sijbers, “Iterative reweighted linear least squares for accurate, fast, and robust estimation of diffusion magnetic resonance parameters: IRLLS for Estimation of Diffusion MR Parameters,” *Magnetic Resonance in Medicine*, vol. 73, no. 6, pp. 2174–2184, Jun. 2015. [Online]. Available: <https://onlinelibrary.wiley.com/doi/10.1002/mrm.25351>
- [26] E. Garyfallidis, M. Brett, B. Amirbekian, A. Rokem, S. Van Der Walt, M. Descoteaux, and I. Nimmo-Smith, “Dipy, a library for the analysis of diffusion MRI data,” *Frontiers in Neuroinformatics*, vol. 8, Feb. 2014, publisher: Frontiers. [Online]. Available: <https://www.frontiersin.org/journals/neuroinformatics/articles/10.3389/fninf.2014.00008/full>
- [27] J. Sjölund, F. Szczepankiewicz, M. Nilsson, D. Topgaard, C.-F. Westin, and H. Knutsson, “Constrained optimization of gradient waveforms for generalized diffusion encoding,” *Journal of Magnetic Resonance (San Diego, Calif.: 1997)*, vol. 261, pp. 157–168, Dec. 2015.
- [28] F. Szczepankiewicz, J. Sjölund, E. Dall’Armellina, S. Plein, J. E. Schneider, I. Teh, and C.-F. Westin, “Motion-compensated gradient waveforms for tensor-valued diffusion encoding by constrained numerical optimization,” *Magnetic Resonance in Medicine*, vol. 85, no. 4, pp. 2117–2126, 2021. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.28551>

- [29] L. Mueller, M. Afzali, S. Coveney, A. Döring, F. Fasano, J. Evans, I. Teh, E. Dall’Armellina, F. Szczepankiewicz, D. K. Jones, and J. E. Schneider, “ZOOM and enhance: ZOnally magnified Oblique Multi-slice for cardiac DTI with ultra-strong gradients,” *Proceedings of the Annual Meeting of ISMRM*, 2024.
- [30] J. Sjölund, “jsjol/NOW,” Sep. 2024, original-date: 2017-03-01T09:06:49Z. [Online]. Available: <https://github.com/jsjol/NOW>
- [31] M. R. Symms, C. A. Wheeler-Kingshott, G. J. M. Parker, and G. J. Barker, “ZOnally-Magnified Oblique Multislice (ZOOM) EPI,” *International Society for Magnetic Resonance in Medicine (ISMRM)*, 2000.
- [32] T. C. Lauenstein, P. Sharma, T. Hughes, K. Heberlein, D. Tudorascu, and D. R. Martin, “Evaluation of optimized inversion-recovery fat-suppression techniques for T2-weighted abdominal MR imaging,” *Journal of Magnetic Resonance Imaging*, vol. 27, no. 6, pp. 1448–1454, 2008. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jmri.21350>
- [33] B. C. Lowekamp, D. T. Chen, L. Ibáñez, and D. Blezek, “The design of simpleitk,” *Frontiers in neuroinformatics*, vol. 7, p. 45, 2013.
- [34] S. Ben-Amitay, D. K. Jones, and Y. Assaf, “Motion correction and registration of high b-value diffusion weighted images,” *Magnetic Resonance in Medicine*, vol. 67, no. 6, pp. 1694–1702, 2012. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.23186>
- [35] R. E. Tracy and G. E. Sander, “Histologically Measured Cardiomyocyte Hypertrophy Correlates with Body Height as Strongly as with Body Mass Index,” *Cardiology Research and Practice*, vol. 2011, p. 658958, Jun. 2011. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3123935/>
- [36] H. Lu, J. H. Jensen, A. Ramani, and J. A. Helpert, “Three-dimensional characterization of non-gaussian water diffusion in humans using diffusion kurtosis imaging,” *NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magnetic Resonance In vivo*, vol. 19, no. 2, pp. 236–247, 2006.