Received <day> <Month>, <year>; | Revised <day> <Month>, <year>; | Accepted <day> <Month>, <year>

DOI: xxx/xxxx

SUPPLEMENTARY FILE

SNR-enhanced diffusion MRI with structure-preserving low-rank denoising in reproducing kernel Hilbert spaces

Gabriel Ramos-Llordén*[1] | Gonzalo Vegas-Sánchez-Ferrero[2] | Congyu Liao[3] | Carl-Fredrik Westin[2] | Kawin Setsompop[3] | Yogesh Rathi[1,2]

[1]Department of Psychiatry, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA
[2]Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA
[3]Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA

Correspondence
*Gabriel Ramos-Llordén. Email: gramosllorden@mgh.harvard.edu

Abstract

This supplementary file includes additional results that complement the content of the main paper. In particular, we include the complete mathematical derivation of KPCA, details on the benefits of the Gaussian kernel, and theoretical insights on the SURE method for optimal parameter design. Finally, additional results for the simulation experiment as well as the in-vivo data experiments are also presented.

KEYWORDS:
diffusion MRI, denoising, noise, PCA, SNR, low-rank, kernel

1 | THEORETICAL RESULTS

1.1 | Mathematical proof of KPCA

In this section, we demonstrate that applying linear PCA in the feature space defined by the mapping $\boldsymbol{\phi}(\cdot)$ can be done just from knowledge of $k(\boldsymbol{x}, \boldsymbol{y}) = \langle \boldsymbol{\phi}(\boldsymbol{x}), \boldsymbol{\phi}(\boldsymbol{y}) \rangle$. Our starting point is a function $\boldsymbol{\phi}(\cdot)$ that maps the diffusion signal $\boldsymbol{x}$ from the native space $\mathbb{R}^M$ to the feature space the $\mathcal{F}$, with $dim(\mathcal{F}) = P$. The mapped noisy signals, $\boldsymbol{y}_n = \boldsymbol{x}_n + \boldsymbol{w}_n$, are denoted as $\boldsymbol{\phi}(\boldsymbol{y}_n)$, $n = 1, ..., N$. As in conventional PCA, we work with the sample covariance $\boldsymbol{S}(P \times P)$ n the feature space $\mathcal{F}$,

$$\boldsymbol{S} = \frac{1}{N} \sum_{n=1}^{N} \tilde{\boldsymbol{\phi}}(\boldsymbol{y}_n) \tilde{\boldsymbol{\phi}}(\boldsymbol{y}_n)^T , \tag{S1}$$

where $\tilde{\boldsymbol{\phi}}(\boldsymbol{y}_n)$ are the centered data , $\tilde{\boldsymbol{\phi}}(\boldsymbol{y}_n) = \boldsymbol{\phi}(\boldsymbol{y}_n) - \bar{\boldsymbol{\phi}}$, with $\bar{\boldsymbol{\phi}} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{\phi}(\boldsymbol{y}_n)$.

The principal component directions $\hat{\boldsymbol{u}}_k$ are obtained from the following eigenvalue problem

$$\boldsymbol{S}\hat{\boldsymbol{u}}_k = \hat{\lambda}_k \hat{\boldsymbol{u}}_k \quad k = 1, \ldots, P . \tag{S2}$$

Rewritting Eq. S2 as

$$\frac{1}{N} \sum_{n=1}^{N} \boldsymbol{\phi}(\boldsymbol{y}_n) \boldsymbol{\phi}(\boldsymbol{y}_n)^T \hat{\boldsymbol{u}}_k = \hat{\lambda}_k \hat{\boldsymbol{u}}_k \tag{S3}$$

we immediately find that the eigenvectors $\hat{\boldsymbol{u}}_k$ can be written as a linear combination of noisy data $\tilde{\boldsymbol{\phi}}(\boldsymbol{y}_n)$, i.e, there exist coefficients α_{kn} such that

$$\hat{\boldsymbol{u}}_k = \sum_{n=1}^{N} \alpha_{kn} \tilde{\boldsymbol{\phi}}(\boldsymbol{y}_n) \quad k = 1, \ldots, P . \tag{S4}$$

By plugging Eq. S4 into Eq. S3, we arrive to

$$\frac{1}{N}\sum_{n=1}^{N}\tilde{\boldsymbol{\phi}}(\boldsymbol{y}_n)\left(\sum_{i=1}^{N}\alpha_{ki}\tilde{\boldsymbol{\phi}}(\boldsymbol{y}_n)^T\tilde{\boldsymbol{\phi}}(\boldsymbol{y}_i)\right)=\hat{\lambda}_k\sum_{i=1}^{N}\alpha_{ki}\tilde{\boldsymbol{\phi}}(\boldsymbol{y}_i)\,. \tag{S5}$$

Multiplying Eq. S5 by $\boldsymbol{\phi}(\boldsymbol{y}_m)^T$ on both left sides, and calling $\tilde{k}(\boldsymbol{y}_m,\boldsymbol{y}_i)=\langle\tilde{\boldsymbol{\phi}}(\boldsymbol{y}_m),\tilde{\boldsymbol{\phi}}(\boldsymbol{y}_i)\rangle=\tilde{\boldsymbol{\phi}}(\boldsymbol{y}_m)^T\tilde{\boldsymbol{\phi}}(\boldsymbol{y}_i)$, we get

$$\frac{1}{N}\sum_{n=1}^{N}\tilde{k}(\boldsymbol{y}_m,\boldsymbol{y}_n)\left(\sum_{i=1}^{N}\alpha_{ki}\tilde{k}(\boldsymbol{y}_n,\boldsymbol{y}_i)\right)=\hat{\lambda}_k\sum_{i=1}^{N}\alpha_{ki}\tilde{k}(\boldsymbol{y}_m,\boldsymbol{y}_i)\,, \tag{S6}$$

or, in matrix form

$$\tilde{\boldsymbol{K}}\tilde{\boldsymbol{K}}\boldsymbol{\alpha}_k=N\hat{\lambda}_k\tilde{\boldsymbol{K}}\alpha_k\,. \tag{S7}$$

If we remove $\tilde{\boldsymbol{K}}$ from both sides ($\tilde{\boldsymbol{K}}$ is invertible), we get the typical eigenvalue problem of KPCA,

$$\tilde{\boldsymbol{K}}\boldsymbol{\alpha}_k=N\hat{\lambda}_k\boldsymbol{\alpha}_k\,, \tag{S8}$$

where $\boldsymbol{\alpha}_k=[\alpha_{k1},\alpha_{k2},\ldots,\alpha_{kN}]^T$ and $\tilde{\boldsymbol{K}}$ the so-called centered kernel matrix, with $\tilde{\boldsymbol{K}}=\boldsymbol{HKH}$, where $\boldsymbol{H}=\boldsymbol{I}-\frac{1}{N}\boldsymbol{1}\boldsymbol{1}^T$ and $\boldsymbol{K}$ is the kernel matrix associated to the original (uncentered) kernel, $\boldsymbol{K}_{mn}=k(\boldsymbol{y}_m,\boldsymbol{y}_n)=\langle\boldsymbol{\phi}(\boldsymbol{y}_m),\boldsymbol{\phi}(\boldsymbol{y}_n)\rangle$. The relation between the two kernels is easy to discern. For any $\boldsymbol{z}$

$$\tilde{k}(\boldsymbol{z},\boldsymbol{y}_n)=\langle\tilde{\boldsymbol{\phi}}(\boldsymbol{z}),\tilde{\boldsymbol{\phi}}(\boldsymbol{y}_n)\rangle=\langle\boldsymbol{\phi}(\boldsymbol{z})-\bar{\boldsymbol{\phi}},\boldsymbol{\phi}(\boldsymbol{y}_n)-\bar{\boldsymbol{\phi}}\rangle=\langle\boldsymbol{\phi}(\boldsymbol{z})\boldsymbol{\phi}(\boldsymbol{y}_n)\rangle-\langle\boldsymbol{\phi}(\boldsymbol{z}),\bar{\boldsymbol{\phi}}\rangle-\langle\bar{\boldsymbol{\phi}},\tilde{\boldsymbol{\phi}}(\boldsymbol{y}_n)\rangle-\langle\bar{\boldsymbol{\phi}},\bar{\boldsymbol{\phi}}\rangle\,, \tag{S9}$$

and if we substitute $\bar{\boldsymbol{\phi}}$ by $\frac{1}{N}\sum_{n=1}^{N}\boldsymbol{\phi}(\boldsymbol{y}_n)$, both kernels are related as

$$\tilde{k}(\boldsymbol{z},\boldsymbol{y}_n)=k(\boldsymbol{z},\boldsymbol{y}_n)-\frac{1}{N}\sum_{i=1}^{N}k(\boldsymbol{z},\boldsymbol{y}_i)-\frac{1}{N}\sum_{i=1}^{N}k(\boldsymbol{y}_i,\boldsymbol{y}_n)+\frac{1}{N^2}\sum_{i,j=1}^{N}k(\boldsymbol{y}_i,\boldsymbol{y}_j)\,. \tag{S10}$$

By solving Eq. S8 , coefficients α_{kn} are identified and principal component directions $\hat{\boldsymbol{u}}_k$ univocally determined by Eq.S4. Since principal component directions $\hat{\boldsymbol{u}}_k$ of Eq. S2 needs to have unit norm, vector $\boldsymbol{\alpha}_k$ needs to be normalized as well, in particular, they are constrained to $N\hat{\lambda}_k||\boldsymbol{\alpha}_k||_2^2=1$.

In summary, given a kernel $k(\cdot,\cdot)$, we calculate matrix $\boldsymbol{K}$, and solve the eigenvalue problem of Eq.S8 . After normalization of $\boldsymbol{\alpha}_k$, the N first principal component directions get determined by plugging coefficients of $\boldsymbol{\alpha}_k$ into Eq. S4. Note that though $k=1,\ldots,\boldsymbol{P}$, we do not have access to all of the $\boldsymbol{P}$ principal component directions but just to N of those, as the eigenvalue problem of Eq. S8 only have N distinct solutions.

We still do not know $\boldsymbol{\phi}(\boldsymbol{y}_n)$, $n=1,\ldots,N$, explicitly. However, the projection of any signal onto the k-th non-linear principal component $\hat{\boldsymbol{u}}_k$ turns out to be computable. In particular, the coefficient of the projection of the target noisy dMRI signal $\boldsymbol{\phi}(\boldsymbol{y}^*)$ onto $\hat{\boldsymbol{u}}_k$ is

$$\beta_k\triangleq\tilde{\boldsymbol{\phi}}(\boldsymbol{y}^*)^T\hat{\boldsymbol{u}}_k=\sum_{n=1}^{N}\alpha_{kn}\tilde{\boldsymbol{\phi}}(\boldsymbol{y}^*)^T\tilde{\boldsymbol{\phi}}(\boldsymbol{y}_n)=\sum_{n=1}^{N}\alpha_{kn}\tilde{k}(\boldsymbol{y}^*,\boldsymbol{y}_n)\,. \tag{S11}$$

Therefore, the projection of the target noisy signal $\boldsymbol{\phi}(\boldsymbol{y}^*)$ onto the first $K_{\mathcal{T}}$ principal component directions (see Eq. 5 of the main body paper) can be given by

$$\hat{\boldsymbol{\phi}}(\boldsymbol{x}^*)=\boldsymbol{P}\boldsymbol{\phi}(\boldsymbol{y}^*)\triangleq\bar{\boldsymbol{\phi}}+\sum_{k=1}^{K_{\mathcal{T}}}\beta_k\hat{\boldsymbol{u}}_k\,. \tag{S12}$$

Let us now focus on the reconstruction problem that gives the denoised signal (Eq. 7 of the main body paper), that is,

$$\hat{\boldsymbol{x}}^*=\arg\min_{\boldsymbol{x}}||\boldsymbol{\phi}(\boldsymbol{x})-\boldsymbol{P}\boldsymbol{\phi}(\boldsymbol{y}^*)||_2^2\,. \tag{S13}$$

We first note that $||\boldsymbol{\phi}(\boldsymbol{x})||_2^2=k(\boldsymbol{x},\boldsymbol{x})$ and $||\boldsymbol{P}\boldsymbol{\phi}(\boldsymbol{y}^*)||_2^2$ is a constant for minimization purposes. Therefore,

$$\hat{\boldsymbol{x}}^*=\arg\min_{\boldsymbol{x}}k(\boldsymbol{x},\boldsymbol{x})-2\boldsymbol{\phi}(\boldsymbol{x})^T\boldsymbol{P}\boldsymbol{\phi}(\boldsymbol{y}^*)\,. \tag{S14}$$

Secondly, by substituting Eq. S4 into Eq. S12 and expressing the centered data $\tilde{\boldsymbol{\phi}}(\boldsymbol{y}_n)$ in terms of $\boldsymbol{\phi}(\boldsymbol{y}_n)$, $\boldsymbol{P}\boldsymbol{\phi}(\boldsymbol{y}^*)$ can be written as

$$\boldsymbol{P}\boldsymbol{\phi}(\boldsymbol{y}^*)=\sum_{n=1}^{N}\gamma_n\boldsymbol{\phi}(\boldsymbol{y}_n)\,, \tag{S15}$$

with

$$\gamma_n = \sum_{k=1}^{K_{\mathcal{F}}} \beta_k \alpha_{kn} + 1/N(1 - \sum_{n=1}^{N} \sum_{k=1}^{K_{\mathcal{F}}} \beta_k \alpha_{kn}) . \tag{S16}$$

Consequently, Eq.S14 tuns out to be

$$\hat{\boldsymbol{x}}^* = \arg\min_{\boldsymbol{x}} k(\boldsymbol{x}, \boldsymbol{x}) - 2 \sum_{n=1}^{N} \gamma_n k(\boldsymbol{x}, \boldsymbol{y}_n) . \tag{S17}$$

1.2 | The feature space defined by the Gaussian kernel is infinite-dimensional

The feature space that the Gaussian kernel is of infinite dimension. This can be demonstrated by means of the MacLaurin series for the exponential function. Indeed,

$$k(\boldsymbol{x}, \boldsymbol{y}) = \langle \boldsymbol{\phi}(\boldsymbol{x}), \boldsymbol{\phi}(\boldsymbol{y}) \rangle = e^{\frac{-||\boldsymbol{x}||_2^2}{2h^2}} e^{\frac{-||\boldsymbol{y}||_2^2}{2h^2}} e^{\frac{\boldsymbol{x}^T \boldsymbol{y}}{h^2}} = \sum_{j=0}^{\infty} \frac{(\boldsymbol{x}^T \boldsymbol{y})^j}{h^{2j} j!} e^{\frac{-||\boldsymbol{x}||_2^2}{2h^2}} e^{\frac{-||\boldsymbol{y}||_2^2}{2h^2}} = \sum_{j=0}^{\infty} \left(\frac{e^{\frac{-||\boldsymbol{x}||_2^2}{2jh^2}}}{h\sqrt{j!}^{1/j}} \frac{e^{\frac{-||\boldsymbol{y}||_2^2}{2jh^2}}}{h\sqrt{j!}^{1/j}} \boldsymbol{x}^T \boldsymbol{y} \right)^j$$

$$= \sum_{j=0}^{\infty} \sum_{\sum_i n_i = j} \frac{e^{\frac{-||\boldsymbol{x}||_2^2}{2jh^2}}}{h\sqrt{j!}^{1/j}} \frac{e^{\frac{-||\boldsymbol{y}||_2^2}{2jh^2}}}{h\sqrt{j!}^{1/j}} \binom{j}{n_1, \dots, n_M} x_1 {y_1}^{n_1} \cdots x_M {y_M}^{n_M} , \tag{S18}$$

where $\boldsymbol{x} = [x_1, \dots, x_M]^T$ and $\boldsymbol{y} = [y_1, \dots, y_M]^T$, and where we have made use of the multinomial theorem for expressing arbitrary powers of $\boldsymbol{x}^T \boldsymbol{y}$. We can hence define the implicit mapping $\boldsymbol{\phi} : \mathbb{R}^M \mapsto \mathcal{F}$ as

$$\boldsymbol{\phi}(\boldsymbol{x}) = \left(\sum_{j=0}^{\infty} \sum_{\sum_i n_i = j} \frac{e^{\frac{-||\boldsymbol{x}||_2^2}{2jh^2}}}{h\sqrt{j!}^{1/j}} \binom{j}{n_1, \dots, n_M}^{1/2} {x_1}^{n_1} \cdots {x_M}^{n_M} \right)_{j=0,\dots,\infty, \sum_{i=1}^{k} n_i = j} , \tag{S19}$$

which proves $k(\boldsymbol{x}, \boldsymbol{y}) = \langle \boldsymbol{\phi}(\boldsymbol{x}), \boldsymbol{\phi}(\boldsymbol{y}) \rangle$ holds. Note that the parameter h controls the shape of the mapping, and that the j-th component decays by a factor $\frac{e^{\frac{-||\boldsymbol{x}||_2^2}{2jh^2}}}{h\sqrt{j!}^{1/j}}$ with increasing h. In fact, it can be proved that for $h \to \infty$, KPCA with $k(\boldsymbol{x}, \boldsymbol{y})$ a Gaussian Kernel function that behaves as linear PCA.

1.3 | Equivalence of Gaussian Kernel PCA and conventional PCA when $h \to \infty$

To prove the equivalence, we will show that the centered Gaussian kernel evaluated at $\boldsymbol{y}_m$ and $\boldsymbol{y}_i$, that is, $\tilde{k}(\boldsymbol{y}_m, \boldsymbol{y}_i) = \tilde{\boldsymbol{\phi}}(\boldsymbol{y}_m)^T \tilde{\boldsymbol{\phi}}(\boldsymbol{y}_i)$, behaves as $\tilde{k}_{\text{linear}}(\boldsymbol{y}_m, \boldsymbol{y}_i) = \frac{1}{h}(\boldsymbol{y}_m - \bar{\boldsymbol{y}})^T (\boldsymbol{y}_i - \bar{\boldsymbol{y}})$ with $\bar{\boldsymbol{y}} = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{y}_i$, when $h \to \infty$. If that is so, we can identify $\boldsymbol{\phi}(\boldsymbol{y}_m) = \frac{1}{\sqrt{h}} \boldsymbol{y}_m$, which is equivalent (up to a constant) to applying PCA over $\boldsymbol{y}_m$ directly.

A first order Taylor expansion of $e^{\frac{-||\boldsymbol{y}_m - \boldsymbol{y}_i||_2^2}{2h^2}}$ with respect of h^2 gives the approximate kernel

$$k(\boldsymbol{y}_m, \boldsymbol{y}_i) \approx 1 - \frac{||\boldsymbol{y}_m - \boldsymbol{y}_i||_2^2}{2h^2} = k_a(\boldsymbol{y}_m, \boldsymbol{y}_i) + k_b(\boldsymbol{y}_m, \boldsymbol{y}_i) , \tag{S20}$$

with $k_a(\boldsymbol{y}_m, \boldsymbol{y}_i) = 1 - \frac{||\boldsymbol{y}_m||_2^2}{2h^2} - \frac{||\boldsymbol{y}_i||_2^2}{2h^2}$ and $k_b(\boldsymbol{y}_m, \boldsymbol{y}_i) = \frac{\boldsymbol{y}_m^T \boldsymbol{y}_i}{h^2}$. The centered kernel $\tilde{k}_a(\boldsymbol{y}_m, \boldsymbol{y}_i)$ (Eq.11 of the main paper applied to $k_a(\boldsymbol{y}_m, \boldsymbol{y}_i)$) gives

$$\tilde{k}_a(\boldsymbol{y}_m, \boldsymbol{y}_i) = 1 - \frac{||\boldsymbol{y}_m||_2^2}{2h^2} - \frac{||\boldsymbol{y}_i||_2^2}{2h^2} - \left(1 - \frac{||\boldsymbol{y}_m||_2^2}{2h^2} - \sum_{i=1}^{N} \frac{||\boldsymbol{y}_i||_2^2}{N2h^2} \right) - \left(1 - \sum_{m=1}^{N} \frac{||\boldsymbol{y}_m||_2^2}{N2h^2} - \frac{||\boldsymbol{y}_i||_2^2}{2h^2} \right)$$

$$+ \left(1 - \sum_{m=1}^{N} \frac{||\boldsymbol{y}_i||_2^2}{N2h^2} - \sum_{i=1}^{N} \frac{||\boldsymbol{y}_i||_2^2}{N2h^2} \right) = 0 . \tag{S21}$$

On the other hand, $\tilde{k}_b(\boldsymbol{y}_m, \boldsymbol{y}_i)$ becomes

$$\tilde{k}_b(\boldsymbol{y}_m, \boldsymbol{y}_i) = \frac{\boldsymbol{y}_m^T \boldsymbol{y}_i}{h^2} - \frac{\boldsymbol{y}_m^T \left(1/N \sum_{i=1}^{N} \boldsymbol{y}_i \right)}{h^2} - \frac{\left(1/N \sum_{m=1}^{N} \boldsymbol{y}_m^T \right) \boldsymbol{y}_i}{h^2} + \frac{1}{N^2} \sum_{m,i=1}^{N} \frac{\boldsymbol{y}_m^T \boldsymbol{y}_i}{h^2} = \frac{1}{h} (\boldsymbol{y}_m - \bar{\boldsymbol{y}})^T (\boldsymbol{y}_i - \bar{\boldsymbol{y}}) . \tag{S22}$$

As a consequence,

$$\tilde{k}(\boldsymbol{y}_m, \boldsymbol{y}_i) \approx \tilde{k}_{\text{linear}}(\boldsymbol{y}_m, \boldsymbol{y}_i) = \frac{1}{h}(\boldsymbol{y}_m - \bar{\boldsymbol{y}})^T(\boldsymbol{y}_i - \bar{\boldsymbol{y}}) \,. \tag{S23}$$

1.4 | Solution of the KPCA denoising problem with Gaussian Kernel

For a Gaussian kernel, $k(\boldsymbol{x}, \boldsymbol{x}) = 1$, and hence Eq. 8 of the main paper simplifies even further,

$$\hat{\boldsymbol{x}}^* = \arg\min_{\boldsymbol{x}} \sum_{n=1}^{N} \gamma_n k(\boldsymbol{x}, \boldsymbol{y}_n) \,. \tag{S24}$$

The first-order condition (gradient equal to zero) for $\hat{\boldsymbol{x}}^*$ to be an extreme point implies [1]

$$\sum_{n=1}^{N} \gamma_n e^{\frac{-||\hat{\boldsymbol{x}}^* - \boldsymbol{y}_n||_2^2}{2h^2}} \left(\hat{\boldsymbol{x}}^* - \boldsymbol{y}_n\right) = 0 \,, \tag{S25}$$

and reordering terms,

$$\hat{\boldsymbol{x}}^* = \frac{\sum_{n=1}^{N} \gamma_n \exp\left(-\frac{||\hat{\boldsymbol{x}}^* - \boldsymbol{y}_n||_2^2}{2h^2}\right) \boldsymbol{y}_n}{\sum_{n=1}^{N} \gamma_n \exp\left(-\frac{||\hat{\boldsymbol{x}}^* - \boldsymbol{y}_n||_2^2}{2h^2}\right)} \,. \tag{S26}$$

In [4], an approximation was made to solve the previous non-linear equation. In particular, $||\hat{\boldsymbol{x}}^* - \boldsymbol{y}_n||_2^2 \approx ||P\boldsymbol{\phi}(\boldsymbol{y}^*) - \boldsymbol{y}_n||_2^2$, in words, the distance between the projected noisy signal $P\boldsymbol{\phi}(\boldsymbol{y}^*)$ to $\boldsymbol{y}_m$ is very similar to that obtained with the optimal solution. With this approximation, $\hat{\boldsymbol{x}}^*$ can be obtained in analytical form as

$$\hat{\boldsymbol{x}}^* = \frac{\sum_{n=1}^{N} \gamma_n \left(1 - 1/2||P\boldsymbol{\phi}(\boldsymbol{y}^*) - \boldsymbol{\phi}(\boldsymbol{y}_n)||_2^2\right) \boldsymbol{y}_n}{\sum_{n=1}^{N} \gamma_n \left(1 - 1/2||P\boldsymbol{\phi}(\boldsymbol{y}^*) - \boldsymbol{\phi}(\boldsymbol{y}_n)||_2^2\right)} \,, \tag{S27}$$

where distances $||P\boldsymbol{\phi}(\boldsymbol{y}^*) - \boldsymbol{\phi}(\boldsymbol{y}_n)||_2^2$ in the feature space can be obtained analytically with formulas given in [2].

1.5 | Stein Unbiased Risk Estimate (SURE)

The Stein's unbiased risk estimate (SURE) provides a means of assessing the true mean square error from the measured data only, without knowledge from the noise-free signal [5].

Given the linear model,

$$\boldsymbol{y}^* = \boldsymbol{x}^* + \boldsymbol{w}, \tag{S28}$$

with $\boldsymbol{w}$ zero-mean uncorrelated Gaussian noise with standard deviation σ, we would like to denoise the signal with optimal values of h and $K_{\mathcal{F}}$, in terms of MSE, $E\{||\boldsymbol{x}^* - \hat{\boldsymbol{x}}^*(h, K_{\mathcal{F}})||_2^2\}$. The SURE associated of $E\{||\boldsymbol{x}^* - \hat{\boldsymbol{x}}^*(h, K_{\mathcal{F}})||_2^2\}$ can be shown to be [5]

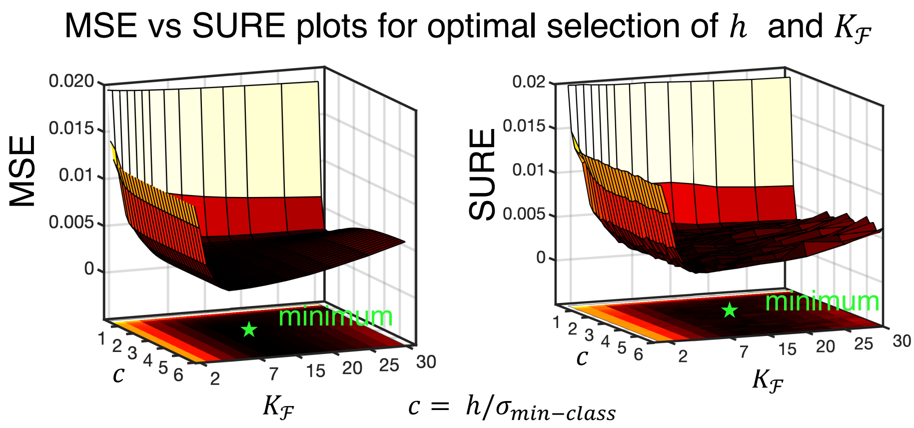
$$\text{SURE}(\hat{\boldsymbol{x}}^*(h, K_{\mathcal{F}})) = ||\boldsymbol{y}^* - \hat{\boldsymbol{x}}^*(h, K_{\mathcal{F}})||_2^2 - M\sigma^2 + 2\sigma^2 \text{div}_{\boldsymbol{y}^*}\{\hat{\boldsymbol{x}}^*(h, K_{\mathcal{F}})\} \,, \tag{S29}$$

where div is the divergence operator of $\hat{\boldsymbol{x}}^*(h, K_{\mathcal{F}})$ with respect to $\boldsymbol{y}^*$. Note that, though we omit it in the notation for sake of clarity, $\hat{\boldsymbol{x}}^*(h, K_{\mathcal{F}})$ is a M-dimensional function of $\boldsymbol{y}^*$. It can be rigorously proved that [6]

$$E\{\text{SURE}(\hat{\boldsymbol{x}}^*(h, K_{\mathcal{F}}))\} = E\{||\boldsymbol{x}^* - \hat{\boldsymbol{x}}^*(h, K_{\mathcal{F}})||_2^2\} \,, \tag{S30}$$

which suggests the use of SURE as a proxy for the MSE. In other words, choosing the optimal parameters that minimizes $\text{SURE}(\hat{\boldsymbol{x}}^*(h, K_{\mathcal{F}}))$ is equivalent (in the expectation sense) to selecting those that make the MSE minimum. See in Figure S1 , the graphs of both the MSE and the SURE are shown as two-dimensional function of h and $K_{\mathcal{F}}$, for one the noisy signals generated in the MC simulation of the main body paper.

Note that when $\sigma \to 0$, SURE converges to the MSE of the noisy and denoised signal, suggesting that, to minimize the MSE, we should not apply any denoising at all, as expected.



Suppporting Information Figure S1 MSE and SURE as two-dimensional functions of rank $K_{\mathcal{F}}$ and mapping parameter h (parameterized by c and $\sigma_{min-class}$. SURE can act as as surrogate for the unobservable MSE in optimal parameter design tasks.

2 | EXPERIMENTAL RESULTS

2.1 | Simulations

Accuracy, precision, and NRMSE results for the dMRI signal are shown in Tables S1 , S2 , and S3 . Results for FA and MD estimation are shown in Tables S4 - S9 .

2.2 | In vivo human brain submillimeter dMRI data

Residuals as well as NRMSE and noise maps for the 860 μm isotropic resolution gSlider data are presented in Figures S2 and S3 , respectively. Fractional anisotropy and fODF-based results are shown in Figures S4 and S5 . Accuracy and precision results for signal estimation, FA, and MD are presented in Tables S10 , and S11 . Absolute error maps for FA and MD are shown in Figures S6 and S7 .

2.2.1 | In vivo human brain conventional low resolution dMRI multishell data

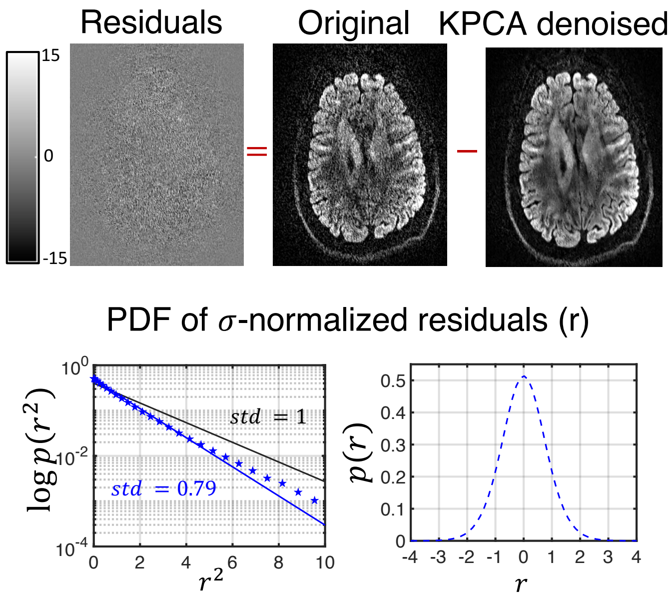
Denoised DWIs for two different shells are shown in Fig. S8 ,whereas corresponding NRMSE (for dMRI signal) and noise maps are display in Fig. S9 . Residual results are shown in Fig. S10 . Accuracy, precision, and NRMSE result for signal estimation, FA, MD and MK are presented in Tables S12 , S13 , and S14 . Absolute error maps are shown in Figure S11 .

2.3 | Capturing non-linear coil and diffusion redundancy simultaneously

Residuals , noise maps, and error maps for the multicoil dMRI data are shown in Figures S12 , S13 and S14 , respectively. NMRSE-based resuls as well as SNR gain are shown in presented (Tab. S15), whereas accuracy and precision results are presented in Table S16 and S17 .

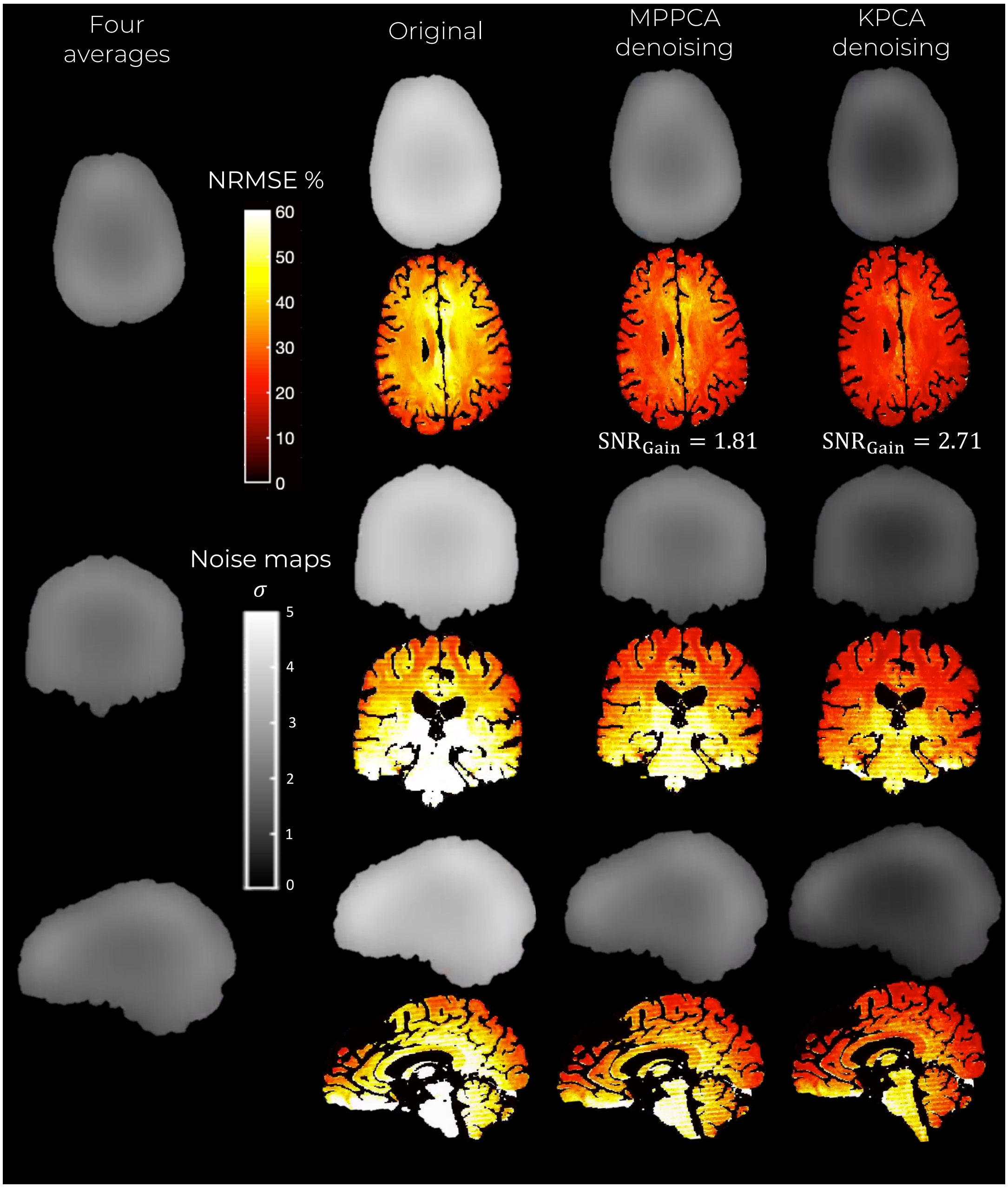
References

1. Mika S, Schölkopf B, Smola AJ, Müller KR, Scholz M, Rätsch G. Kernel PCA and de-noising in feature spaces. *Adv Neural Inf Process Syst*, 1999; 536–542.



Suppporting Information Figure S2 Residual maps from the 860 μm resolution datasets after being denoised with KPCA. On top of the figure, the residual map from a given DW image, which shows no anatomical information. On the bottom, the probability density function of the residuals (r) normalized by the level of noise, σ. For the statistics, the normalized residuals are taken for all diffusion directions and number of repetitions. Observe that the residuals for KPCA approximately follows a Gaussian distribution (blue dotted line on both plots representing the estimated pdf). On blue solid-line the optimal analytical zero-mean Gaussian distribution that best fits the data (Maximum Likelihood sense). Note that the standard deviation of the normalized residual, 0.79, is lower than 1 (black-line represents a zero-mean standard Gaussian distribution).

2. Kwok JY, Tsang IH. The pre-image problem in kernel methods. *IEEE Trans Neural Netw Learn Syst* 2004; **15**(6):1517–1525.
3. Welch BL. The generalization ofstudent's' problem when several different population variances are involved. *Biometrika* 1947; **34**(1/2):28–35.
4. Rathi Y, Dambreville S, Tannenbaum A. Statistical shape analysis using kernel pca. *Image processing: algorithms and systems, neural networks, and machine learning*, vol. 6064, International Society for Optics and Photonics, 2006; 60 641B.
5. Ramani S, Blu T, Unser M. Monte-carlo sure: A black-box optimization of regularization parameters for general denoising algorithms. *IEEE Trans Image Process* 2008; **17**(9):1540–1554.
6. Eldar YC. Generalized SURE for exponential families: Applications to regularization. *IEEE Trans Signal Process* 2008; **57**(2):471–481.



Suppporting Information Figure S3 Maps of the NRMSE (hot colormap) and noise level (gray colormap) for the denoised DW images at 860 μm isotropic resolution and b-value of $b = 2000\ s/\text{mm}^2$. Observe that KPCA denoising obtains lowest level of noise (highest SNR gains) and NRMSE.

		Gray Matter									White Matter								
		$b = 1200\ s/\mathrm{mm}^2$			$b = 1500\ s/\mathrm{mm}^2$			$b = 2500\ s/\mathrm{mm}^2$			$b = 1200\ s/\mathrm{mm}^2$			$b = 1500\ s/\mathrm{mm}^2$			$b = 2500\ s/\mathrm{mm}^2$		
		Directions (M)			Directions (M)			Directions (M)			Directions (M)			Directions (M)			Directions (M)		
	SNR	32	64	128	32	64	128	32	64	128	32	64	128	32	64	128	32	64	128
Original		0.5	0.6	0.5	0.8	0.8	0.7	1.7	1.7	1.6	0.4	0.6	0.6	0.8	0.9	0.8	2.4	1.9	1.9
MMPCA	5	7.6	7.4	6.3	9.6	9.3	7.9	17.7	16.4	15.8	4	2.7	2.5	5.1	4.1	4	19.7	12.9	13.3
KPCA		7.3	7.3	7.3	9.3	9.3	9.2	16.7	16.9	16.9	7.4	11	9.7	9.2	14.4	12.6	22.8	34.1	30
Original		0.4	0.4	0.4	0.5	0.5	0.5	0.8	1.1	1.1	0.4	0.4	0.4	0.4	0.5	0.5	1.4	1.1	1.3
MMPCA	8	6.3	4.4	3.6	8.1	5.5	4.6	15.5	13.9	10.3	2.4	2.5	2.5	3.8	4.1	3.9	12.6	11	6.5
KPCA		6.5*	6.3*	5.9*	8.2*	7.95*	7.5*	15.4*	15.1*	14.8*	4.1*	3.6*	3.2*	5.3*	4.6*	4.3*	12.9*	11.2*	10.7*
Original		0.3	0.3	0.3	0.4	0.3	0.4	0.8	0.8	0.8	0.3	0.3	0.3	0.4	0.5	0.4	0.9	1.1	1
MPPCA	15	4.5	3.6	3.1	5.8	4.6	4	14.2	9.9	8.2	2.4	2.5	2.4	3.9	3.9	3	11.4	6.3	4.7
KPCA		5.9	5.6	5.3	7.5	7	6.6	14.4	13.8	13.1	3.7	3.3	2.9	4.7	4.2	4	11.6	10.3	9.7

Suppporting Information Table S1 Absolute bias [%] of the original, MPCA-denoised, and KPCA- denoised dMRI signals, compared to the ground truth dMRI signal. MC-based simulation experiment. Differences in cases marked with * are not statistically significant as confirmed with a Welch's t-test (p-value $< .001$)[3].

		Gray Matter									White Matter								
		$b = 1200\ s/\mathrm{mm}^2$			$b = 1500\ s/\mathrm{mm}^2$			$b = 2500\ s/\mathrm{mm}^2$			$b = 1200\ s/\mathrm{mm}^2$			$b = 1500\ s/\mathrm{mm}^2$			$b = 2500\ s/\mathrm{mm}^2$		
		Directions (M)			Directions (M)			Directions (M)			Directions (M)			Directions (M)			Directions (M)		
	SNR	32	64	128	32	64	128	32	64	128	32	64	128	32	64	128	32	64	128
Original		52.6	52.6	52.5	67	66.9	66.9	151.6	151.1	151.3	55	55.1	55.5	71.7	72.3	72.7	182.1	182.2	185
MMPCA	5	11.9	10.7	11.5	15.2	13.4	14.4	27	29.1	25.8	25.7	21.2	18	33.9	27.9	23.6	84	70.1	61.3
KPCA		9.6	7.6	6.4	12.1	9.6	8.1	26.2	20.6	17.3	16.4	10.9	9.8	21.4	14.1	12.9	52.7	34.3	31.9
Original		32.9	32.8	32.9	42	42	42	95	94	95	34.4	34.4	34.7	44.9	45	45.5	114	113	116
MMPCA	8	10	10.5	9.8	12.5	13.2	12.5	22.1	22.9	24.7	17	13.2	11	22.1	17.3	14.8	55.6	48.8	45.2
KPCA		6.8*	5.5*	4.7*	8.6*	7.1*	6*	17.9*	14.5*	12.2*	9.8*	8.3*	7.5*	12.5*	10.7*	9.7*	31.4*	27.1*	24.6*
Original		26.3	26.3	26.3	33.5	33.5	33.5	75.9	75.6	75.6	27.6	27.5	27.7	3.6	35.9	36.3	91.2	90.9	92.3
MPPCA	5	9.9	9.2	8.1	12.4	11.7	10.3	21	23.2	22.2	13.4	10.4	8.8	17.6	14.2	13.1	47.2	42.9	37.8
KPCA		5.8	4.8	4.1	7.4	6.1	5.3	15.3	12.5	10.6	7.4	6.4	5.9	9.5	8.3	7.6	23.8	20.9	19.3

Suppporting Information Table S2 Standard deviation [%] of the original, MPCA-denoised, and KPCA- denoised dMRI signals, compared to the ground truth dMRI signal. MC-based simulation experiment. Differences in cases marked with * are not statistically significant as confirmed with a Welch's t-test (p-value $< .001$)[3].

		Gray Matter									White Matter								
		$b = 1200\ s/\mathrm{mm}^2$			$b = 1500\ s/\mathrm{mm}^2$			$b = 2500\ s/\mathrm{mm}^2$			$b = 1200\ s/\mathrm{mm}^2$			$b = 1500\ s/\mathrm{mm}^2$			$b = 2500\ s/\mathrm{mm}^2$		
		Directions (M)			Directions (M)			Directions (M)			Directions (M)			Directions (M)			Directions (M)		
	SNR	32	64	128	32	64	128	32	64	128	32	64	128	32	64	128	32	64	128
Original		52.6	52.5	52.4	67.0	67.0	67.0	152	151	151	55.0	55.1	55.5	71.7	72.3	72.7	182	182	185
MMPCA	5	14.8	13.6	13.6	18.9	17.0	17.1	33.5	34.7	31.5	26.1	21.5	18.3	34.5	28.5	24.1	86.9	71.8	63.3
KPCA		12.6	11.1	10.2	15.9	14.1	12.9	32.2	27.7	25.3	18.7	16.1	14.4	24.3	20.9	18.8	59.6	49.9	45.1
Original		32.9	32.8	32.9	41.9	41.8	41.8	94.9	94.4	94.5	34.4	34.4	34.7	44.9	44.9	45.5	114	114	116
MMPCA	8	12.3	11.7	10.7	15.5	14.9	13.6	28.5	27.9	27.7	17.3	13.5	11.2	22.6	17.9	15.4	57.6	50.5	45.9
KPCA		9.9*	8.8*	8.0*	12.6*	11.2*	10.2*	24.7*	21.9*	20.1*	11.2*	9.6*	8.7*	14.2*	12.3*	11.3*	35.3*	30.6*	28.1*
Original		17.5	17.5	17.5	22.3	22.3	22.3	50.5	50.5	50.5	18.3	18.4	18.5	23.9	24	24.2	60.8	60.9	61.5
MPPCA	15	8.7	7.4	6.2	11.0	9.5	8.0	22.9	20.2	18.1	9.4	8.1	7.3	13.1	11.7	10.1	37.7	31.7	26.3
KPCA		7.1	6.1	5.3	8.9	7.7	6.6	17.7	15.7	14.1	6.2	5.7	5.1	8.0	6.9	6.8	19.6	16.7	16.2

Suppporting Information Table S3 NRMSE [%] of the original, MPCA-denoised, and KPCA- denoised dMRI signals, compared to the ground truth dMRI signal. MC-based simulation experiment.Differences in cases marked with * are not statistically significant as confirmed with a Welch's t-test p-value $< .001)$[3].

		Gray Matter									White Matter								
		$b = 1200\ s/\text{mm}^2$			$b = 1500\ s/\text{mm}^2$			$b = 2500\ s/\text{mm}^2$			$b = 1200\ s/\text{mm}^2$			$b = 1500\ s/\text{mm}^2$			$b = 2500\ s/\text{mm}^2$		
		Directions (M)			Directions (M)			Directions (M)			Directions (M)			Directions (M)			Directions (M)		
	SNR	32	64	128	32	64	128	32	64	128	32	64	128	32	64	128	32	64	128
Original		126	76.3	46.1	111	65.3	36.13	55.8	17.1	7.3	13.8	8.5	5.8	10.5	5.7	3.5	6.2	10.3	11.9
MMPCA	5	55.2	59. 5	44.5	53.8	59.5	45.2	55.9	52.2	59.8	11.7	7.4	4.9	16.5	11.4	7.8	19.6	16.4	14.2
KPCA		13	7.9	17.8	19.75	4.1	15.2	38.2	7.2	10.5	2.6	2.6	2	5.7	4.6	3.8	14.2	12.3	11.9
Original		69.1	42.6	27.3	79.2	48.5	31.6	57.3	24.3	7.5	9.7	6.5	5.1	10.8	8.1	6.4	1.5	0.9	2.2
MMPCA	8	41.8	16.3	7.1	43.9	16. 6	7.5	55.3	43.3	19.1	6.5	4.1	2.9	9.1	5.9	4.2	18.2	14.5	10.5
KPCA		4.4	9.5	10.1	3.9	8.5	10.1	20.6	3.2	4.2	2.7	2.6	2.1	1.5	1.4	1.6	8.9	7.8	6.7
Original		43.8	25.7	15.9	52.9	34.3	22.1	56.2	30.6	17.5	6.1	4.2	3.4	8.5	6.5	5.5	3.9	2.1	1.3
MPPCA	15	16.1	5.9	3.9	17.4	5.3	2.95	47.3	15.5	3.8	4.8	3.1	2.5	6.4	4.5	3.1	16.3	11.3	8.1
KPCA		7.7	9.8	10.2	6.6	9.2	9.3	8.2	1.8	5.5	3.5	3.3	3	2.5	2.4	2.6	4.6	3.9	3.3

Suppporting Information Table S4 Absolute bias [%] of the FA estimates (compared to ground-truth FA) obtained after LLS estimation of the diffusion tensor from the original dMRI signal and denoised signals with MPPCA and KPCA. MC-based simulation experiment. Differences in cases marked with * are not statistically significant as confirmed with a Welch's t-test (p-value $< .001$)[3].

		Gray Matter									White Matter								
		$b = 1200\ s/\mathrm{mm}^2$			$b = 1500\ s/\mathrm{mm}^2$			$b = 2500\ s/\mathrm{mm}^2$			$b = 1200\ s/\mathrm{mm}^2$			$b = 1500\ s/\mathrm{mm}^2$			$b = 2500\ s/\mathrm{mm}^2$		
		Directions (M)			Directions (M)			Directions (M)			Directions (M)			Directions (M)			Directions (M)		
		32	64	128	32	64	128	32	64	128	32	64	128	32	64	128	32	64	128
	SNR																		
Original		83.8	64.9	49.1	87.7	66.9	49.4	105	74.6	53.5	19.61	13.6	9.8	20.1	14.4	9.7	26.8	16.6	11.6
MMPCA	5	30	31.2	36.5	31.8	32.5	37.7	40.9	37.1	31.5	17	12.2	8.5	18.1	13.5	9	15.8	10.9	8.2
KPCA		40.5	32.9	28.7	45.4	35.3	30.6	62.6	45.4	37.1	8.9	7.1	4.7	9.8	7.5	4.9	12.5	8.6	5.8
Original		53.5	39.4	28.9	59.9	44.2	32.3	74.3	56.1	40.8	12.4	8.6	6.2	12.8	8.8	6.1	16.4	10.9	7.3
MMPCA	8	43.7	38.7	28.3	43.6	38.9	29.3	36.6	45.8	42.2	10	7.1	5.1	11	7.4	5.2	11.9	9	7.4
KPCA		30.1	25.9	21.1	31.3	26.1	21.5	50.5	38.6	31.9	4	3.1	2.4	3.9	3	2.2	8.2	5.8	4
Original		40.7	30.1	21.9	45.7	34.8	24.2	67.4	49.1	35.5	9.6	6.7	4.7	10.6	7.3	5	13.3	8.9	6
MPPCA	15	42.4	31.3	22.8	42.6	31.9	23	45.6	46.3	35.6	7.9	5.6	3.9	8.3	5.7	4.1	10.9	8.9	7.1
KPCA		26.3	21.4	16.3	26.5	21.3	16.5	40.7	32	25.3	2.9	2.1	1.6	2.9	2.1	1.5	5.9	4.1	2.7

Suppporting Information Table S5 Standard deviation [%] of the FA estimates (compared to ground-truth FA) obtained after LLS estimation of the diffusion tensor from the original dMRI signal and denoised signals with MPPCA and KPCA. MC-based simulation experiment. Differences in cases marked with * are not statistically significant as confirmed with a Welch's t-test (p-value $< .001$)[3].

		Gray Matter									White Matter								
		$b = 1200\ s/mm^2$			$b = 1500\ s/mm^2$			$b = 2500\ s/mm^2$			$b = 1200\ s/mm^2$			$b = 1500\ s/mm^2$			$b = 2500\ s/mm^2$		
		Directions (M)			Directions (M)			Directions (M)			Directions (M)			Directions (M)			Directions (M)		
	SNR	32	64	128	32	64	128	32	64	128	32	64	128	32	64	128	32	64	128
Original		152	100	67.3	142	93.5	61.2	119	76.6	54.0	23.9	16.0	11.4	22.7	15.4	10.3	27.5	19.6	16.6
MMPCA	5	62.8	67.1	57.6	62.6	67.8	58.9	69.3	64.1	67.6	20.7	14.3	9.8	24.6	17.7	11.9	25.2	19.7	16.5
KPCA		42.5	33.9	33.8	49.5	35.6	34.2	73.3	45.9	38.6	9.2	7.6	5.2	11.4	8. 8	6.2	18.9	15.1	13.3
Original		87.4	58.0	39.8	99.3	65.6	45.2	93.8	61.1	41.5	15.7	10.8	8.0	16.8	11.9	8.9	16.5	11.0	7.6
MMPCA	8	60.5	41.9	29.2	61.9	42.4	30.2	66.4	63.0	46.3	12.0	8.2	5.8	14.3	9.4	6.6	21.7	17.0	12.8
KPCA		30.4	27.6	23.3	31.5	27.5	23.8	54.5	38.7	32.2	4.8	4.0	3.2	4.3	3.4	2.7	12.1	9.7	7.8
Original		32.7	22.2	15.2	36.2	24.4	16.3	71.2	49.2	35.3	6.1	4.3	3.1	7.5	5.5	4.1	11.7	8.6	6.6
MPPCA	15	28.6	20.9	14.5	29.3	21.2	14.8	42.9	30.8	22.4	5.9	4.3	2.9	6.6	4.4	3.1	13.9	9.9	6.7
KPCA		20.8	19.6	13.6	20.9	18.8	11.7	26.1	20.5	16.0	4.3	4.1	2.1	3.8	1.9	2.0	2.9	3.4	3.6

Suppporting Information Table S6 NRMSE [%] of the FA estimates (compared to ground-truth FA) obtained after LLS estimation of the diffusion tensor from the original dMRI signal and denoised signals with MPPCA and KPCA. MC-based simulation experiment. Differences in cases marked with * are not statistically significant as confirmed with a Welch's t-test (p-value $< .001$)[3].

		Gray Matter									White Matter								
		$b = 1200\ s/\text{mm}^2$			$b = 1500\ s/\text{mm}^2$			$b = 2500\ s/\text{mm}^2$			$b = 1200\ s/\text{mm}^2$			$b = 1500\ s/\text{mm}^2$			$b = 2500\ s/\text{mm}^2$		
		Directions (M)			Directions (M)			Directions (M)			Directions (M)			Directions (M)			Directions (M)		
	SNR	32	64	128	32	64	128	32	64	128	32	64	128	32	64	128	32	64	128
Original		18.9	19	19.2	18.9	19	19.1	0.8	0.9	0.8	18	18.5	18.4	14.9	15.4	15.3	7.4	7.1	7.3
MMPCA	5	1.6	1.3	1.1	1.5	1.2	1.4	0.6	2.5	2.1	6.8	5.1	2.9	9.4	7.1	4.3	3.3	6.6	7.7
KPCA		3.5	2.8	2.3	4.7	3.5	2.9	10.1	9.3	8.2	2.9	3.8	3.9	3.9	4.6	4.5*	5.3	7	8.4
Original		7.4	7.3	7.3	10.3	10	10.2	10.2	10.3	10.2	9.2	8.9	9.1	11.7	11.6	11.6	2.2	2.4	2.3
MMPCA	8	1.1	1.1	1.3	1.2	1	1.4	1.8	1.5	1.9	2.7	1.4	1	3.9	2.4	1.7	8.2	8.2	7.2
KPCA		1.8	1.5	1.3*	2	1.5	1.4*	5.8	4.6	3.5	1.6	1.8	0.6	1.4	1.6	1.8*	4.5	4.6	4.5
Original		4.1	4.1	4.3	6.1	6.1	6.1	11.3	11.3	11.4	5.4	5.5	5.5	8.1	8.1	8.1	5.1	5.2	5.1
MPPCA	15	0.9	1.1	1.1	1.1	1.2	1.15	1.5	1.8	2.2	1.5	1	0.8	2.4	1.7	1.2	8.3	7.1	5.9
KPCA		1.4*	1.1	1.2	1.5	1.2	1.2	3.6	2.8	2.3	1.4	1.6	1.2	1.1	1.3	1.4	2.2	2.3	2.4

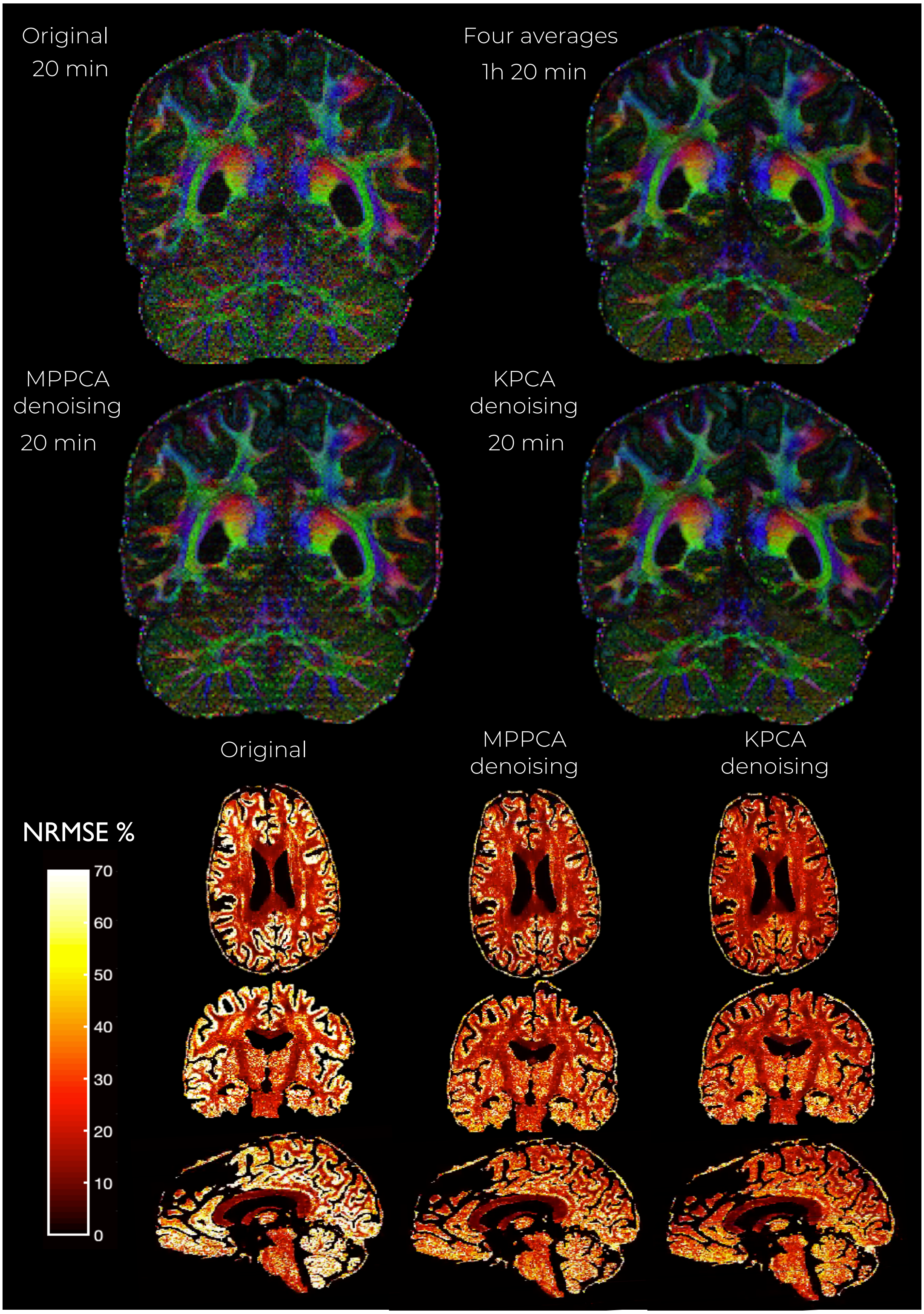
Suppporting Information Table S7 Absolute bias [%] of the MD estimates (compared to ground-truth MD) obtained after LLS estimation of the diffusion tensor from the original dMRI signal and denoised signals with MPPCA and KPCA. MC-based simulation experiment. Differences in cases marked with * are not statistically significant as confirmed with a Welch's t-test (p-value $< .001$)[3].

		Gray Matter									White Matter								
		$b = 1200\ s/\text{mm}^2$			$b = 1500\ s/\text{mm}^2$			$b = 2500\ s/\text{mm}^2$			$b = 1200\ s/\text{mm}^2$			$b = 1500\ s/\text{mm}^2$			$b = 2500\ s/\text{mm}^2$		
		Directions (M)			Directions (M)			Directions (M)			Directions (M)			Directions (M)			Directions (M)		
	SNR	32	64	128	32	64	128	32	64	128	32	64	128	32	64	128	32	64	128
Original		14.1	10.2	7.2	13.5	9.6	6.8	9.7	6.8	4.8	14.5	10.3	7.3	13.1	9.4	6.6	8.9	6.5	4.6
MMPCA	5	9.4	7.1	5.1	9.8	7.1	5.3	10.2	11.2	8.1	9.9	7.4	5.9	11.5	9.1	7.1	8.8	7.5	7.0
KPCA		8.9	7	5.2	9.9	7.3	5.4	12.7	10.7	8.5	5.9	4.2	2.8	6.8	4.8	3.1*	8.7	6.6	4.8
Original		7.8	5.5	3.9	9.1	6.3	4.5	9.1	6.5	4.6	9.4	6.4	4.6	10.1	6.9	4.9	8.4	5.9	4.2
MMPCA	8	5.9	4.4	3.1	6.2	4.6	3.2	8.9	6.5	5.1	7.2	5.1	3.6	8.3	5.6	3.9	9.4	8.7	6.7
KPCA		5.9	4.5	3.1*	6.1	4.7	3.3*	9.6	7.2	5.1	2.8	2.1	1.9	2.9	2.1	1.5*	5.9	4.5	3.2
Original		5.6	4.0	2.7	6.3	4.6	3.2	8.6	5.9	4.3	6.7	4.8	3.3	8.1	5.7	3.9	8.1	5.7	3.9
MPPCA	15	4.8	3.5	2.45	4.9	3.6	2.5	7.2	5.4	3.9	5.6	4	2.8	6.4	4.4	3	10	7.9	5.8
KPCA		4.9*	3.6	2.5	5	3.7	2.6	7.5	5.4	3.8	2	1.4	1	2.1	1.5	1	4.2	3	2

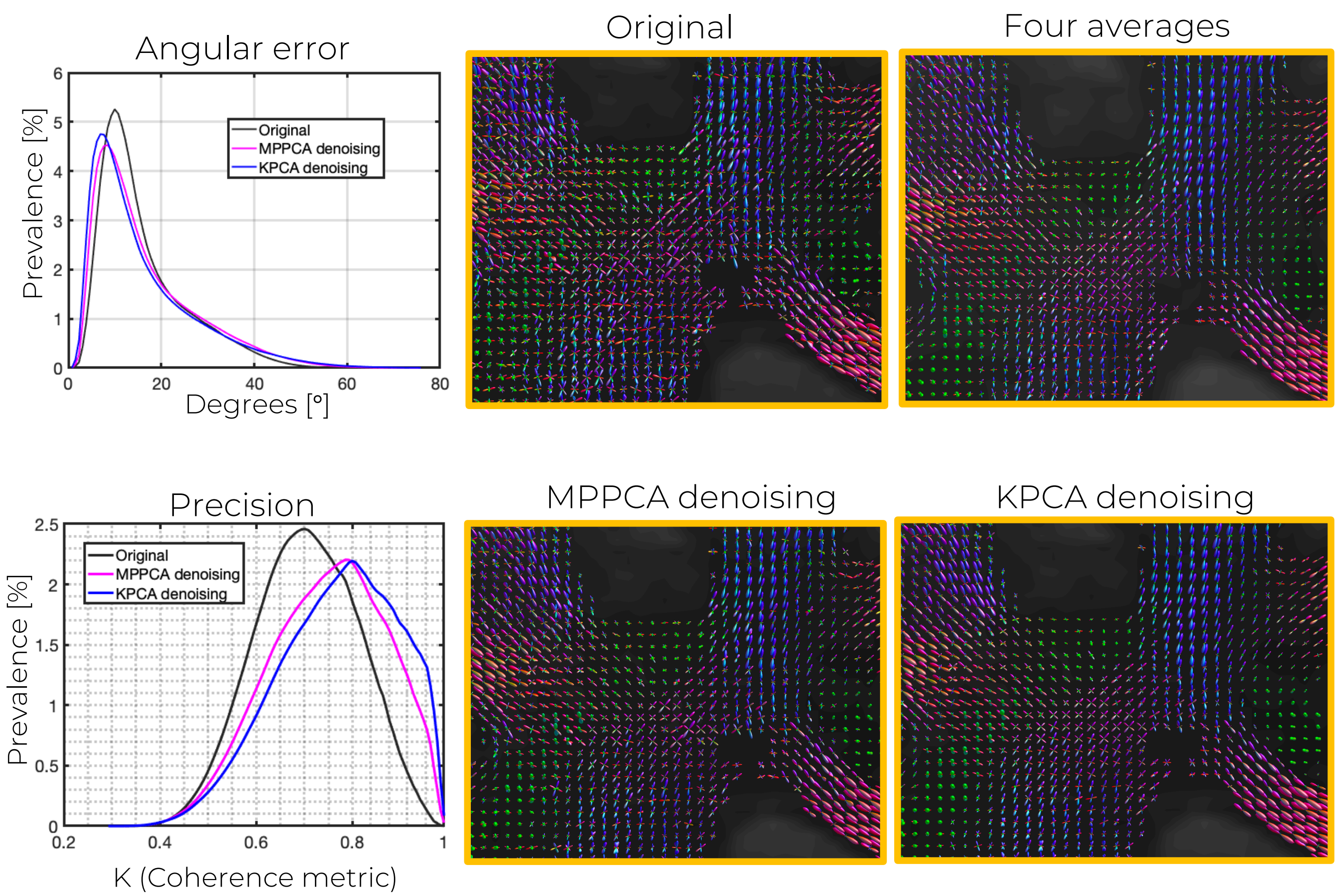
Suppporting Information Table S8 Standard deviation [%] of the MD estimates (compared to ground-truth MD) obtained after LLS estimation of the diffusion tensor from the original dMRI signal and denoised signals with MPPCA and KPCA. MC-based simulation experiment. Differences in cases marked with * are not statistically significant as confirmed with a Welch's t-test (p-value $< .001$)[3].

		Gray Matter									White Matter								
		$b = 1200\ s/\mathrm{mm}^2$			$b = 1500\ s/\mathrm{mm}^2$			$b = 2500\ s/\mathrm{mm}^2$			$b = 1200\ s/\mathrm{mm}^2$			$b = 1500\ s/\mathrm{mm}^2$			$b = 2500\ s/\mathrm{mm}^2$		
		Directions (M)			Directions (M)			Directions (M)			Directions (M)			Directions (M)			Directions (M)		
		32	64	128	32	64	128	32	64	128	32	64	128	32	64	128	32	64	128
	SNR																		
Original		23.6	21.6	20.5	23.2	21.3	20.3	9.7	6.9	4.9	23.2	21.2	19.8	19.8	18.0	16.6	11.6	19.7	8.6
MMPCA	5	9.6	7.2	5.2	9.9	7.3	5.4	10.2	11.5	8.4	12.0	8.9	9.8	6.6	14.9	11.5	8.3	9.4	10.0
KPCA		10.4	9.6	7.6	5.7	11.0	8.1	6.2	16.2	14.1	11.8	6.6	5.7	4.8	7.8	6.7*	5.5	10.2	9.6
Original		10.7	9.2	8.3	13.7	11.8	11.1	13.7	12.1	11.2	13.2	11.0	10.2	15.4	13.6	12.6	8.7	6.4	4.8
MMPCA	8	6.0	4.5	3.4	6.3	4.7	3.5	9.1	6.6	5.4	7.6	5.3	3.7	9.2	6.1	4.3	12.5	11.9	9.8
KPCA		6.3	4.7	3.4*	6.5	4.9	3.6*	11.2	8.5	6.2	3.2	2.8	1.9	3.2	2.7	2.3*	7.4	6.5	5.5
Original		3.9	2.9	2.4	4.3	3.4	2.9	10.7	9.5	9.0	4.3	3.4	2.8	5.7	4.7	4.1	10.3	8.7	8.1
MPPCA	15	3.5	2.4	1.6	3.6	2.5	1.7	5.3	3.8	2.7	3.8	2.6	1.8	4.2	2.9	1.9	10.0	7.2	4.9
KPCA		3.2*	2.3	1.6	3.2	2.3	1.5	5.1	3.6	2.7	1.7	1.6	2.1	1.5	2.8	1.9	2.0	3.2	3.5

Suppporting Information Table S9 NRMSE [%] of the MD estimates (compared to the ground-truth MD) obtained after LLS estimation of the diffusion tensor from the original dMRI signal and denoised signals with MPPCA and KPCA. MC-based simulation experiment. Differences in cases marked with * are not statistically significant as confirmed with a Welch's t-test (p-value $< .001$)[3].



Suppporting Information Figure S4 Color-encoded FA maps of the denoised DW images at 860 μm isotropic resolution and b-value of $b = 2000\ s/\text{mm}^2$ as well as corresponding NRMSE maps.



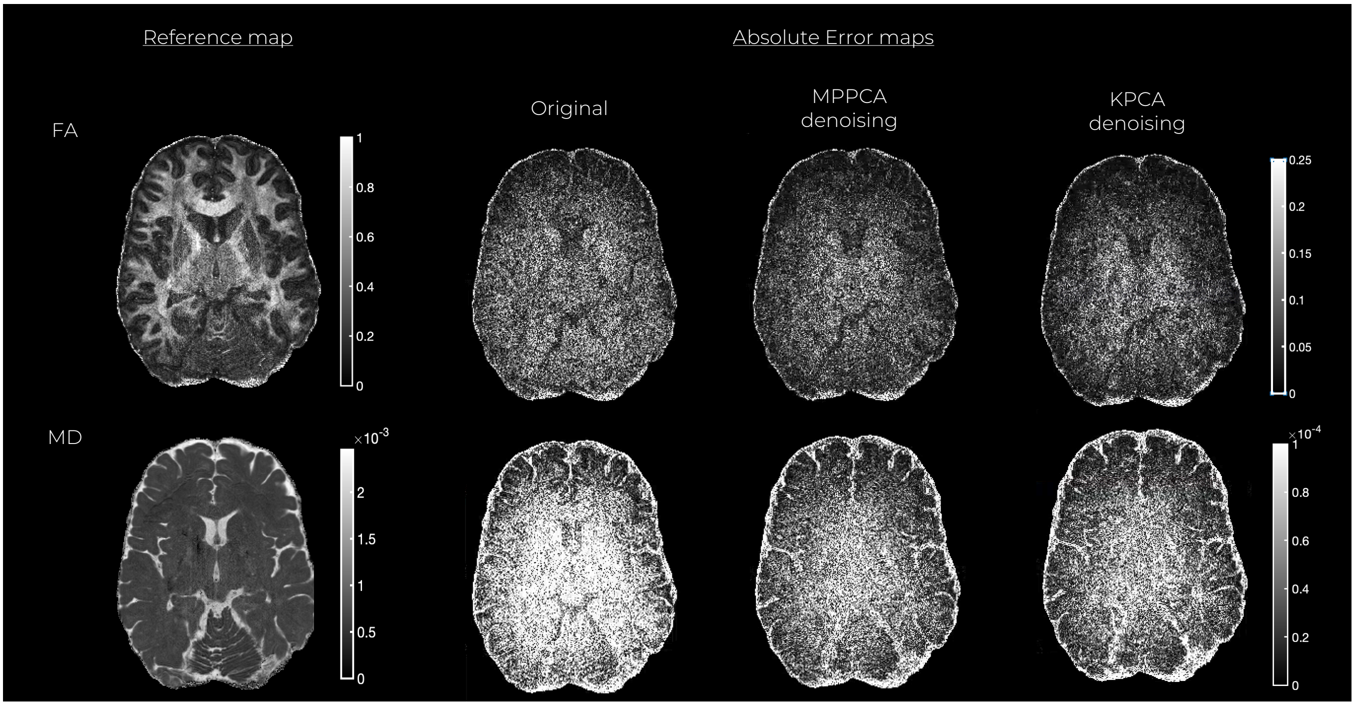
Suppporting Information Figure S5 Angular error as well as angular precision, probed by coherence metric κ, for the peaks of the fODFS estimated with CSD after denoising the 860 μm isotropic resolution DW images ($b = 2000\ s/\mathrm{mm}^2$) . Further, corresponding fODFs maps in a representative crossing-fibers area are displayed. Observe the lower variability in the fODFs of KPCA denoising compared to MPPCA.

	Signal (Bias [%])			FA (Bias [%])			MD (Bias [%])		
	Brain	WM	GM	Brain	WM	GM	Brain	WM	GM
Original-660μm	28	26	26	32	19	48	11	13	7
MPPCA-660μm	23	22	21	19	12	25	3	3	1
KPCA-660μm	19	18	17	17	9	19	3	4	2
Original-860μm	29	28	27	38	27	51	13	15	11
MPPCA-860μm	24	23	23	21	17	25	5	7	2
KPCA-860μm	17	16	16	18	16	20	6	8	2

Suppporting Information Table S10 Accuracy results for the experiment with in-vivo human brain submillimeter resolution dMRI data.

	Signal (Standard deviation [%])			FA (Standard deviation [%])			MD (Standard deviation [%])		
	Brain	WM	GM	Brain	WM	GM	Brain	WM	GM
Original-660μm	27	28	27	32	19	48	11	13	7
MPPCA-660μm	23	21	21	19	12	25	3	3	1
KPCA-660μm	17	16	16	17	38	19	3	4	2
Original-860μm	26	26	25	34	29	41	16	21	9
MPPCA-860μm	20	19	19	32	28	37	15	21	9
KPCA-860μm	16	15	16	27	25	30	14	19	8

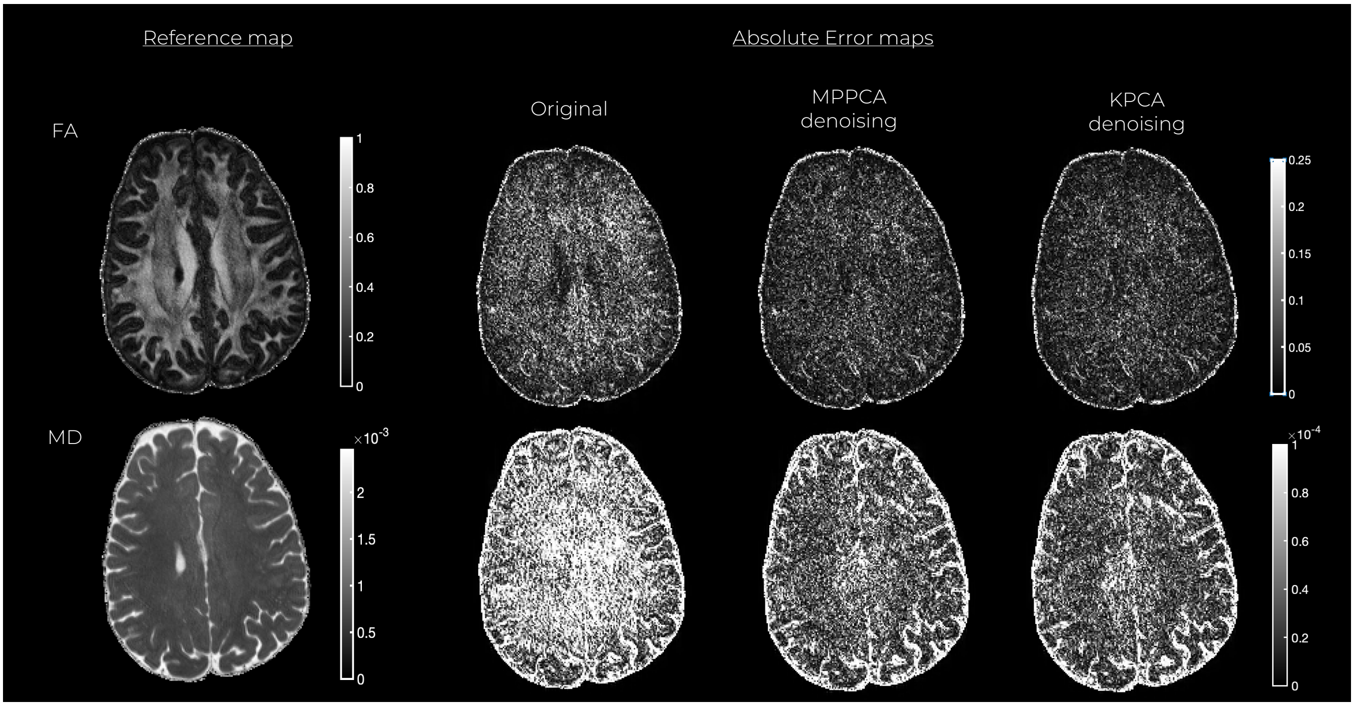
Suppporting Information Table S11 Precision results for the experiment with in-vivo human brain submillimeter resolution dMRI data.



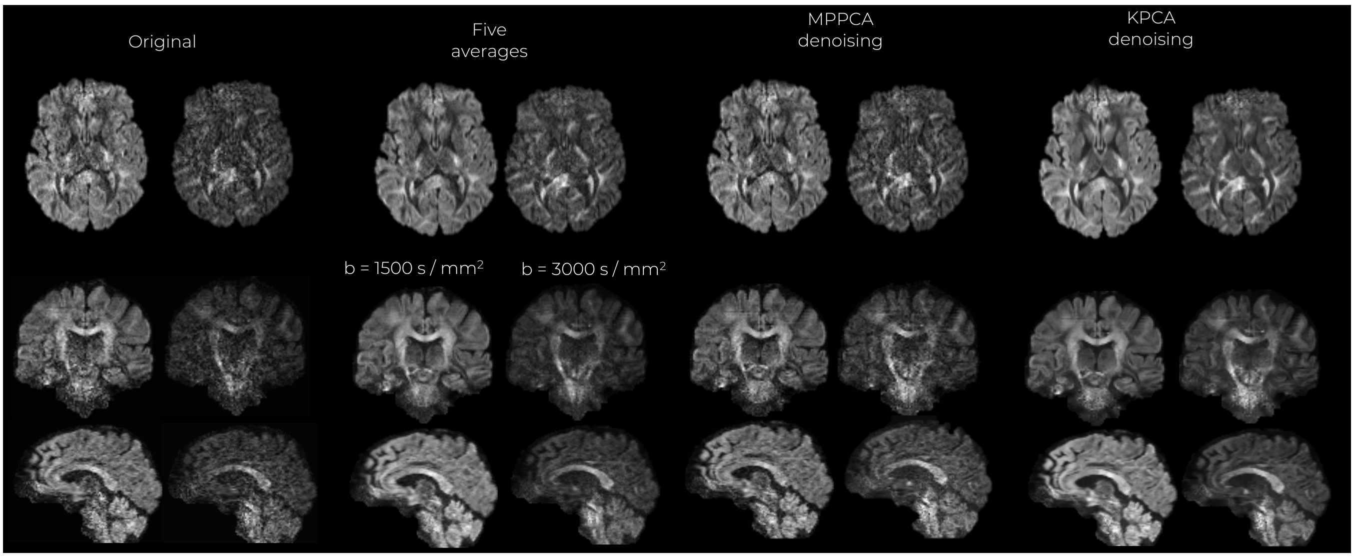
Suppporting Information Figure S6 Absolute error maps of FA and MD for the 660 μm gSlider dataset.

	Signal (NRMSE)			SNRgain	FA (NRMSE)			MD (NRMSE)			MK (NRMSE)		
	Brain	WM	GM	Brain	Brain	WM	GM	Brain	WM	GM	Brain	WM	GM
Original	23%	23%	23%	1x	64%	18%	37 %	12%	8%	9%	18%	11%	23%
MPPCA	20%	20%	20%	1.5 x	43%	14%	28%	12%	7%	9%	15%	8%	19%
KPCA	15%	16%	16%	2.8 x	32 %	12%	24%	11%	6%	8%	13%	6%	16%

Suppporting Information Table S12 NRMSE and SNR-based results from the experiment with conventional multi-shell dMRI data. Note that in all cases KPCA denoising achieves better results than MPPCA.



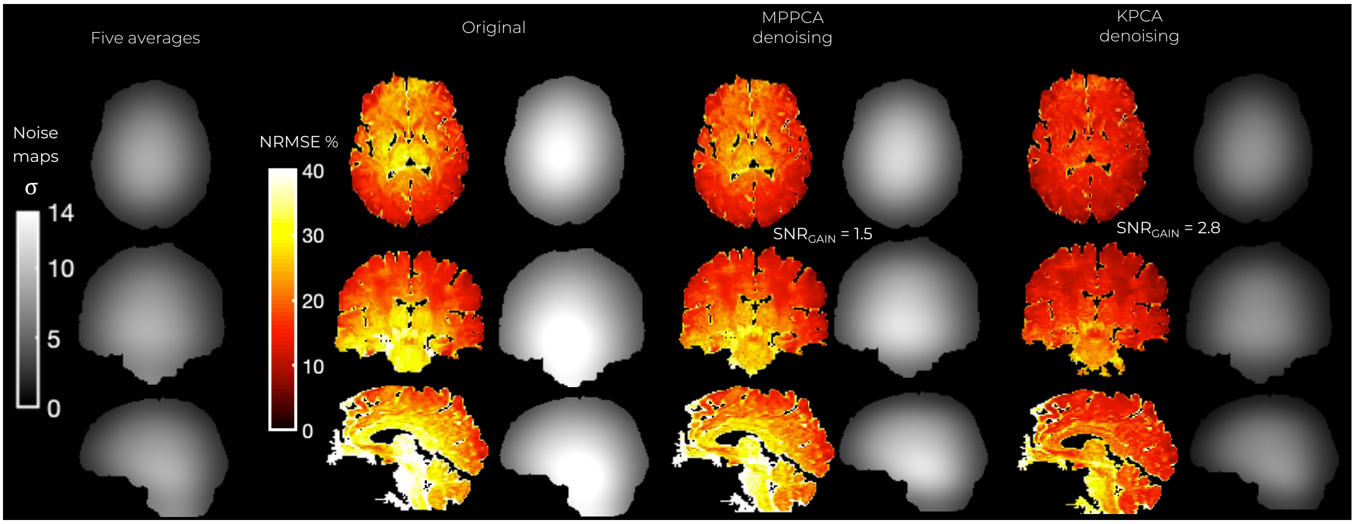
Suppporting Information Figure S7 Absolute error maps of FA and MD for the 860 μm gSlider dataset.



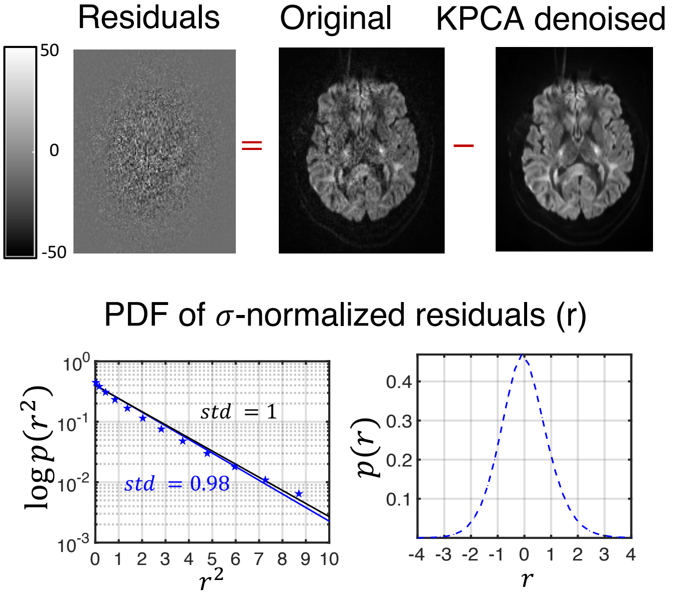
Suppporting Information Figure S8 Mid-axial, coronal and sagittal slices of denoised multi-shell conventional DWI images.

	Signal (Bias [%])			FA (Bias [%])			MD (Bias [%])			MK (Bias [%])		
	Brain	WM	GM	Brain	WM	GM	Brain	WM	GM	Brain	WM	GM
Original	19	16	15	52	18	37	12	7	8	18	9	18
MPPCA	17	14	13	35	14	28	11	6	7	15	7	15
KPCA	13	10	10	27	12	24	11	5	7	13	5	12

Suppporting Information Table S13 Accuracy results for the experiment with conventional multi-shell dMRI data.



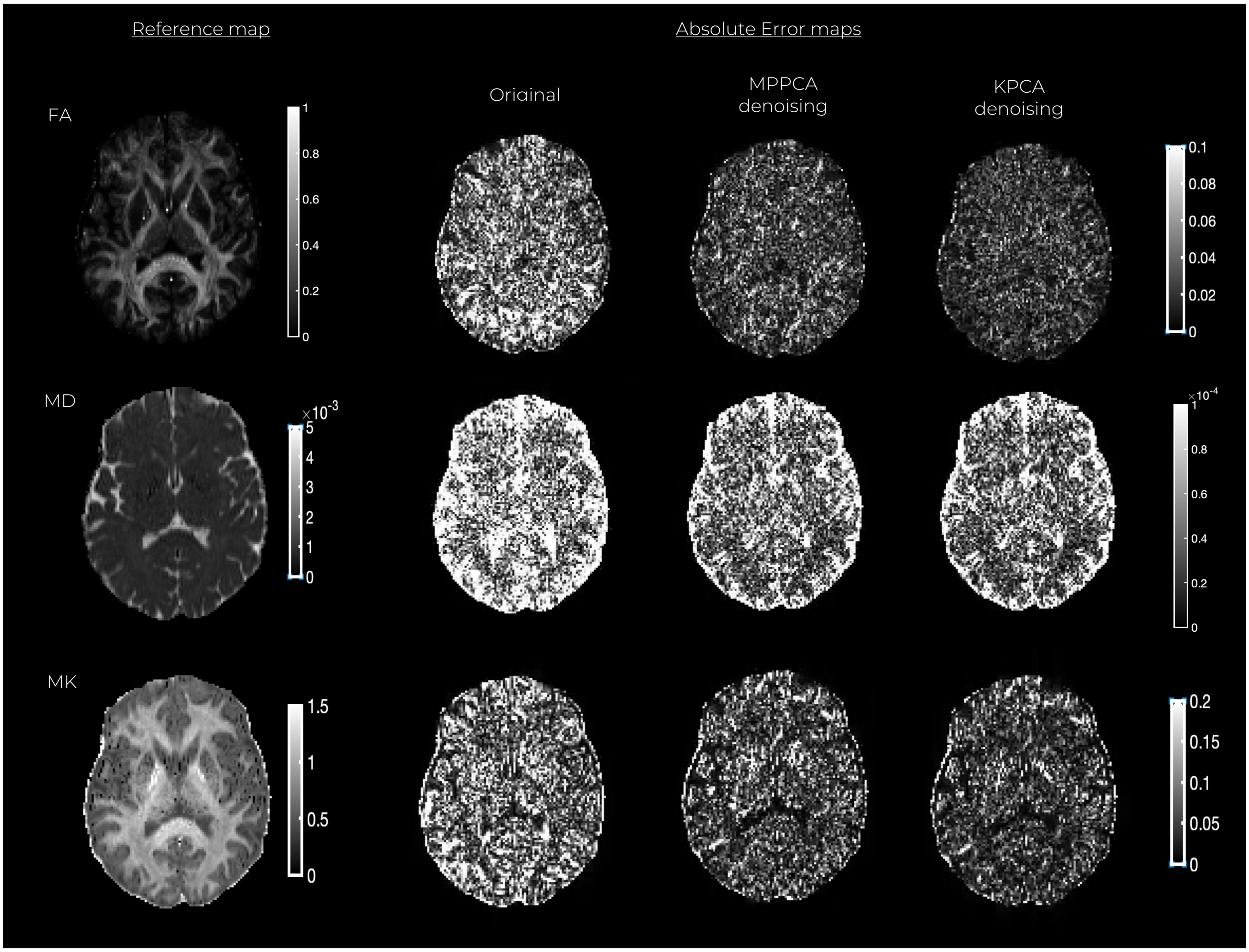
Suppporting Information Figure S9 Maps of the NRMSE (hot colormap) and noise level (gray colormap) for the denoised multi-shell conventional DW images. Observe that KPCA denoising obtains the lowest level of noise (highest SNR gains) and NRMSE.



Suppporting Information Figure S10 Residual maps from the conventional mutil-shell dMRI datasets after being denoised with KPCA. On top of the figure, the residual map from a given DW image, which shows no anatomical information. On the bottom, the probability density function of the residuals (r) normalized by the level of noise, σ. For the statistics, the normalized residuals are taken for all diffusion directions and number of repetitions. Observe that the residuals for KPCA approximately follows a Gaussian distribution (blue dotted line on both plots representing the estimated pdf). On blue solid-line the optimal analytical zero-mean Gaussian distribution that best fits the data (Maximum Likelihood sense). Note that the standard deviation of the normalized residual, 0.98, is lower than 1 (black-line represents a zero-mean standard Gaussian distribution).

	Signal (Standard deviation [%])			FA (Standard deviation[%])			MD (Standard deviation [%])			MK (Standard deviation [%])		
	Brain	WM	GM	Brain	WM	GM	Brain	WM	GM	Brain	WM	GM
Original	22	18	18	46	18	32	12	9	10	18	11	22
MPPCA	17	14	14	37	14	27	12	8	10	15	8	18
KPCA	10	8	8	26	12	22	11	6	9	13	6	15

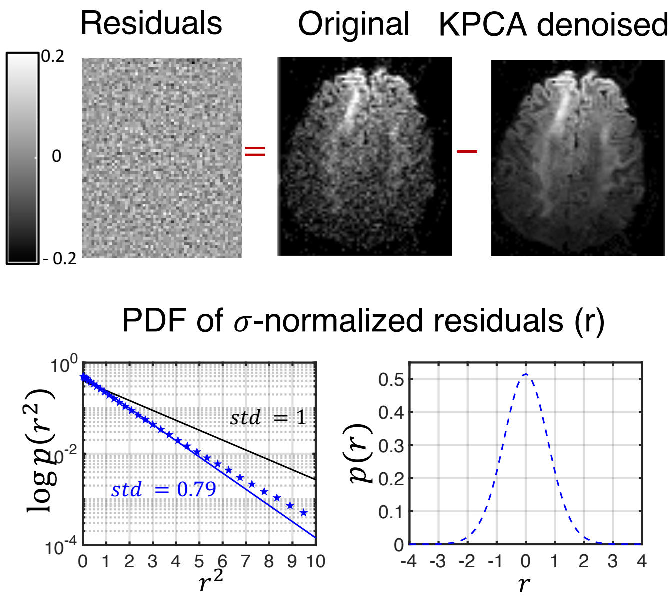
Suppporting Information Table S14 Precision results for the experiment with conventional multi-shell dMRI data.



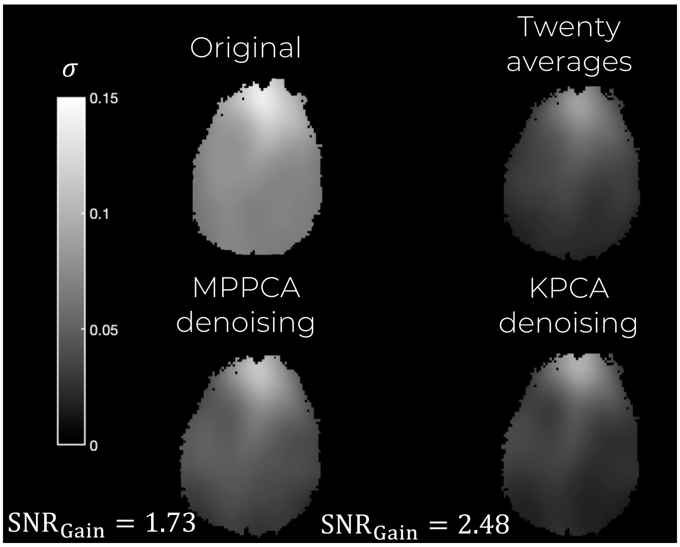
Suppporting Information Figure S11 Absolute error maps of FA, MD, and MK for the conventional low-resolution dMRI dataset.

	Signal (NRMSE)			SNRgain	FA (NRMSE)			MD (NRMSE)		
	Brain	WM	GM	Brain	Brain	WM	GM	Brain	WM	GM
Original	68%	74%	72%	1x	111%	33%	112%	8%	9%	11%
MPPCA	40%	45%	43%	1.73x	67%	27%	72%	8%	8%	10%
KPCA	32%	31%	35%	2.48x	55%	24%	60%	7%	7%	9%

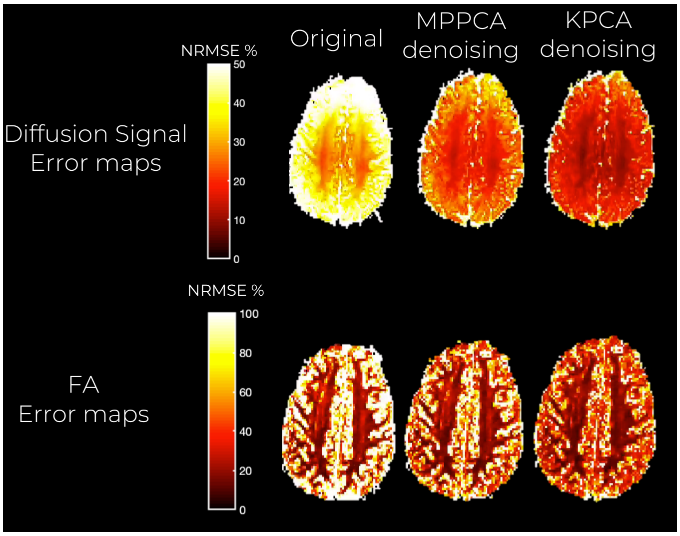
Suppporting Information Table S15 NRMSE and SNR-based results from the experiment with multi-coil dMRI data. Note that in all cases KPCA denoising achieves better results than MPPCA.



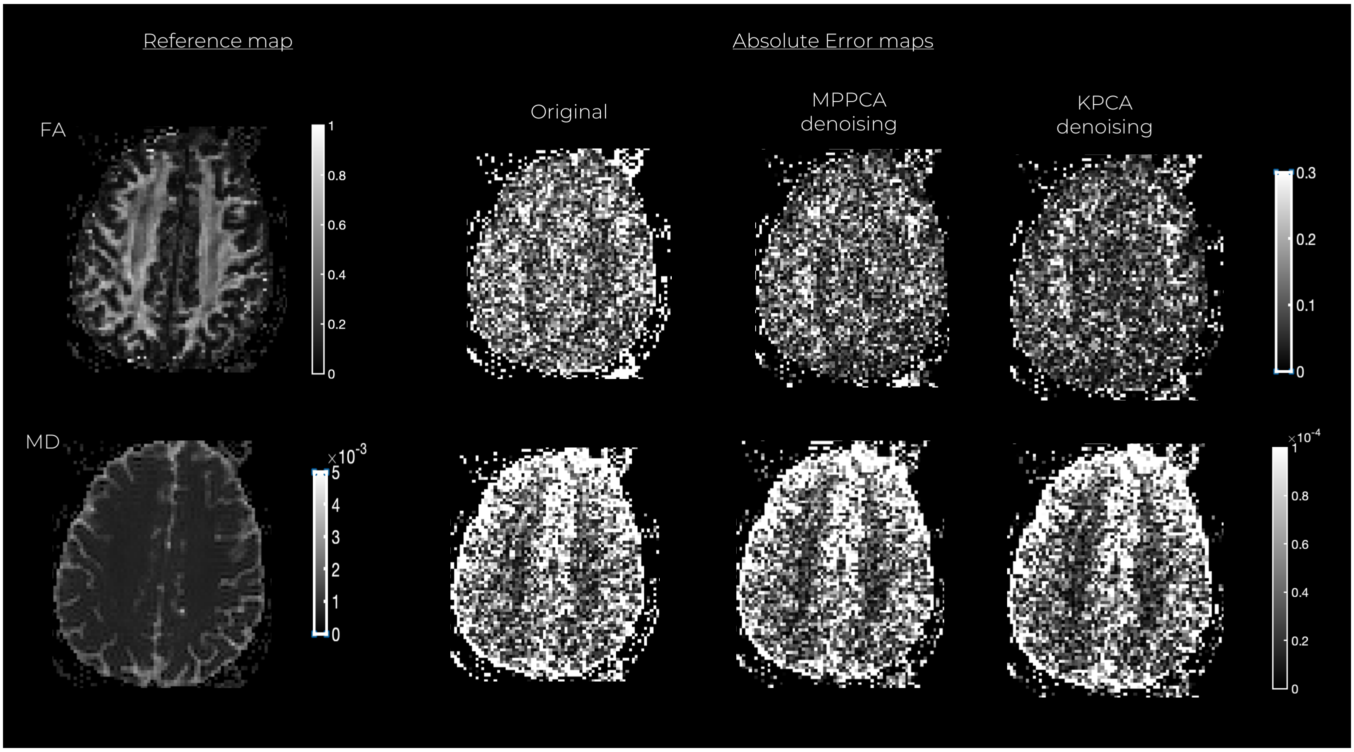
Suppporting Information Figure S12 Residual maps from the multi-coil dMRI datasets after being denoised with KPCA. On top of the figure, the residual map from a given DW image, which shows no anatomical information. On the bottom, the probability density function of the residuals (r) normalized by the level of noise, σ. For the statistics, the normalized residuals are taken for all diffusion directions and number of repetitions. Observe that the residuals for KPCA approximately follows a Gaussian distribution (blue dotted line on both plots representing the estimated pdf). On blue solid-line the optimal analytical zero-mean Gaussian distribution that best fits the data (Maximum Likelihood sense). Note that the standard deviation of the normalized residual, 0.79, is lower than 1 (black-line represents a zero-mean standard Gaussian distribution).



Suppporting Information Figure S13 Estimated noise maps for the denoised multi-coil DW images. Observe that KPCA denoising obtains the lowest level of noise (highest SNR gains).



Suppporting Information Figure S14 NRMSE maps for the denoised multi-coil DW images. Maps of errors for the signal and the fractional anisotropy are shown Observe that KPCA denoising obtains the lowest NRMSE in both type of maps.



Suppporting Information Figure S15 Absolute error maps of FA and MD for the multi-coil dMRI dataset.

	Signal (Bias [%])			FA (Bias [%])			MD (Bias [%])		
	Brain	WM	GM	Brain	WM	GM	Brain	WM	GM
Original	32	34	29	85	19	118	2	2	3
MPPCA	28	29	25	46	10	56	1	1	1
KPCA	22	23	20	35	10	46	1	1	2

Suppporting Information Table S16 Accuracy results for the experiment with multi-coil dMRI data.

	Signal (Standard deviation [%])			FA (Standard deviation [%])			MD (Standard deviation [%])		
	Brain	WM	GM	Brain	WM	GM	Brain	WM	GM
Original-660μm	34	33	33	57	25	66	8	8	11
MPPCA-660μm	26	27	25	49	24	58	8	7	10
KPCA-660μm	21	22	21	43	22	53	7	7	9

Suppporting Information Table S17 Precision results for the experiment with multi-coil dMRI data.