Complex diffusion-weighted image estimation via matrix recovery under general noise models

Lucilio Cordero-Grande [*], Daan Christiaens, Jana Hutter, Anthony N. Price, Jo V. Hajnal

Centre for the Developing Brain and Biomedical Engineering Department, School of Biomedical Engineering and Imaging Sciences, King's College London, King's Health Partners, St Thomas' Hospital, London, SE1 7EH, UK

ABSTRACT

We propose a patch-based singular value shrinkage method for diffusion magnetic resonance image estimation targeted at low signal to noise ratio and accelerated acquisitions. It operates on the complex data resulting from a sensitivity encoding reconstruction, where asymptotically optimal signal recovery guarantees can be attained by modeling the noise propagation in the reconstruction and subsequently simulating or calculating the limit singular value spectrum. Simple strategies are presented to deal with phase inconsistencies and optimize patch construction. The pertinence of our contributions is quantitatively validated on synthetic data, an in vivo adult example, and challenging neonatal and fetal cohorts. Our methodology is compared with related approaches, which generally operate on magnitude-only data and use data-based noise level estimation and singular value truncation. Visual examples are provided to illustrate effectiveness in generating denoised and debiased diffusion estimates with well preserved spatial and diffusion detail.

1. Introduction

Signal to noise ratio (SNR) is the ultimate limiting factor for high resolution magnetic resonance imaging (MRI) (Edelstein et al., 1986; Fuderer, 1988). Diffusion weighted imaging (DWI), aiming at probing the microstructure as given by the diffusivity of water molecules, here with a focus on brain applications, is particularly SNR limited, which is due to the exponential decay of the water signal with the diffusion sensitization factor b. In addition, accelerated encodings have been proposed for denser diffusion sampling, which generally reduce the SNR per volume but increase the SNR per unit time (Setsompop et al., 2012). This has motivated an extensive literature on filtering methods for DWI that have evolved from standard filtering techniques on a per-volume basis (Basu et al., 2006) to more comprehensive procedures that jointly consider the information sampled in the diffusion space (Tristán-Vega and Aja-Fernández, 2010). In parallel, different spatial and diffusion representations have been suggested for DWI, which could aid denoising methods based on prior models. Recently, principal component analysis (PCA) based methods have been proposed for DWI denoising (Pai et al., 2011; Manjón et al., 2013). These techniques are based on the assumption that sampling has been made redundant enough so that the covariance of the signal when arranged in a matrix of local spatial patches of M elements times the number of sampled volumes N can be explained by a number of components R much lower than $\min(M, N)$, with the remaining observed covariance being attributable to noise. Further contributions have proposed to estimate R by resorting to random matrix theory, namely to the Marčenko-Pastur law that governs the empirical spectral noise distribution, both for DWI noise estimation (Veraart et al., 2016a) and denoising (Veraart et al., 2016b). Assuming that the number of components is much lower than the matrix dimensions and the signal is strong enough, the noise will be confined into an interval of well separated small eigenvalues, so that a truncation threshold can be defined from a statistical viewpoint.

Most of the denoising techniques for DWI have been applied to magnitude-only datasets provided by scanner reconstructions, the most widely accessible image data type. However, in magnitude reconstructions the noise statistics depend on the signal level, especially at low SNR. This is typically addressed by post-processing or iterative techniques for Rician bias correction (Koay and Basser, 2006) but these may only provide moderate benefits, particularly when considering the price of significantly increased modeling and computing complexity. Following a different path that recognizes the complex nature of the acquired data simplifies the statistical treatment, particularly in very low SNR regimes, potentially the most rewarding situation when denoising.

---

* Corresponding author.
E-mail address: lucilio.cordero_grande@kcl.ac.uk (L. Cordero-Grande).

This was demonstrated a long time ago (Bernstein et al., 1989; Wood and Johnson, 1999) and applied to wavelet-based DWI denoising on a per volume basis in (Wirestam et al., 2006). Additional studies have investigated the negative influence of the bias introduced by magnitude data in the reliability of derived diffusion features (Jones and Basser, 2004; Bernd et al., 2009).

By founding the denoising criterion in universal properties of noise and being model-free, random matrix based approaches are attractive tools for denoising the complex raw data right after the reconstruction. At this stage, the statistical noise properties can be well characterized and, when conveniently arranged, the signal may admit a compact representation by singular value decomposition (SVD). In contrast, model-based approaches, unless strongly founded in diffusion data properties, may be more prone to suppress relevant information or to introduce spurious features, with magnified risks of altering the diffusion estimates considering usual needs for subsequent data pre-processing. Convincing results were provided in (Veraart et al., 2016b), for instance, about the preservation of high resolution data features after filtering.

Consequently, we propose an extension of the random matrix denoising Marčenko-Pastur PCA (MPPCA) method in (Veraart et al., 2016b). This extension departs from homoscedastic and uncorrelated noise and hard separability of signal and noise assumptions of MPPCA by making use of the results in (Benaych-Georges and Nadakuditi, 2012) for optimal singular value shrinkage under a general Marčenko-Pastur law. Usage of these results allows our method to be applied in general acquisition settings not covered by the independent and identically distributed noise model, such as partial Fourier and temporally alternating encodings, and to replace the truncation operation with optimal shrinkage rules stemming from the noise model under consideration. Thus, we refer to our proposal as generalized singular value shrinkage (GSVS) denoising. Our pipeline preserves noise additivity by operating in the complex domain, with a strategy being introduced to limit the signal complexity due to unpredictable temporal phase variations. In addition, random matrix theory results are fully exploited by making use of a criterion to determine optimal patch sizes (§ 2). In vivo experiments are performed on high *b*-value adult data, a challenging dataset comprised of highly accelerated neonatal DWI acquired with a particularly involved encoding strategy to maximize data robustness against sporadic motion and distortions, and strongly motion degraded low SNR fetal DWI acquisitions. Benefits and/or increased flexibility of modeling the noise properties by accounting for the adopted reconstruction operations, working with complex data while correcting for linear phase inconsistencies, modelling the spectral noise properties, and using singular value shrinkage versus singular value truncation are quantitatively assessed by resorting to estimation risks provided by the random matrix theory. Moreover, we validate our approach using simulations and added value in retrieved diffusion information and derived measures is visually illustrated (§ 3). The source code of a MATLAB implementation of the signal recovery method and the data required to reproduce some of the results in Figs. 4–6 and 9 is made available at https://github.com/mriphysics/complexSVDShrinkageDWI/releases/tag/1.1.0. Finally, we discuss the strengths and limitations of the proposed approach and provide some lines for improvement and extension (§ 4) to end up with some concluding remarks (§ 5).

2. DWI estimation as a matrix recovery problem

DWI estimation in the presence of noise can be formulated as a set of matrix recovery subproblems by arranging the DWI data into a set of matrices $\mathbf{Y}$ of size $M \times N$. Each of these matrices is assumed to be the result of the additive corruption of an underlying (approximately) low but unknown rank R matrix $\mathbf{X}$ with a random matrix with Gaussian noise entries $\mathbf{W}$, so we can write $\mathbf{Y} = \mathbf{X} + \mathbf{W}$. The objective of each subproblem is to provide an estimate of the signal matrix $\widehat{\mathbf{X}}$ with some guarantees on the estimation risk induced by a given loss function $L(\mathbf{X}, \widehat{\mathbf{X}})$ that

quantifies the error between the estimated and the actual data. In what follows we examine how the DWI data can comply with the assumptions of this model and describe the proposed image estimation techniques. First we pay attention to the nature of the complex diffusion signal (§ 2.1), then we describe the propagation of noise in the reconstruction (§ 2.2), later on we present the adopted algorithmic pipeline (§ 2.3), and finally we describe the procedure for matrix recovery (§ 2.4).

2.1. Low-rank DWI signal

For the low-rank condition $R \ll \min(M, N)$ to reasonably hold, we should arrange the DWI data in a way that maximizes the redundancy in the entries of the matrices $\mathbf{Y}$. Arranging the data from local patches along the rows and the diffusion measures along the columns of $\mathbf{Y}$, simultaneously taking advantage of spatial and diffusion redundancy, has been proposed for instance in (Pai et al., 2011; Manjón et al., 2013; Veraart et al., 2016b) as the local PCA model for diffusion denoising. We can define a set of local patch extraction matrices indexed by their center voxel index q, $\mathbf{P}^{[q]}$, $q \in \mathscr{C} = \{1, ..., Q\}$ of size $M \times Q$, with Q the number of voxels in each data volume, $\mathscr{C}$ the set of voxel indexes, and M the number of voxels in the local 3D patch, which are indexed by $m \in \mathscr{M} = \{1, ..., M\}$. We use spherical instead of more common cubic patches, with potential benefits from rotational invariance. Thus, we make $p_{m \leftarrow q_2}^{[q_1]} = 1$ if the voxel indexed by q_2 is the m-closest to the voxel indexed by q_1, and 0 otherwise.

A potential source of spurious complexity in the data is given by artifacts. For complex DWI data, we have to pay special attention to the phase of the signal. The DWI phase varies in a largely unpredictable manner during the acquisition mainly due to patient motion during the scan (Bammer et al., 2010). Larger gradient strengths imply increased levels of phase fluctuations and reduced SNR, so accurate phase corrections become especially difficult for high *b*-values. Thus, our method performs phase corrections on the basis of a robust gross approximation of the underlying phase (see § 2.3), which can increase the redundancy of the signal without introducing undesired biases. For now, we denote the reconstructed volumes for the set of diffusion measurements as $\mathbf{y}$, their phase corrected versions as $\tilde{\mathbf{y}}$, the local information for the patch centered at $[q]$ after phase correction as $\tilde{\mathbf{y}}^{[q]} = \mathbf{P}^{[q]} \tilde{\mathbf{y}}$ and express patch-to-matrix assignments as $\mathbf{Y}^{[q]} \leftarrow \tilde{\mathbf{y}}^{[q]}$. In this setting, DWI denoising involves solving $Q' \leq Q$ matrix denoising subproblems that provide a set of estimates for the signal at q, with one estimate for each patch the voxel q belongs to, so that a patch combination rule can be implemented to build the resulting denoised DWI volumes (see § 2.4).

2.2. Additive Gaussian noise in DWI reconstruction

DWI measures are assumed to follow a circularly symmetric complex Gaussian stationary noise distribution of zero mean and a given variance σ^2, $\mathscr{CN}(0, \sigma^2)$ modeling the Johnson-Nyquist noise observed in the quadrature receiver. In parallel imaging, levels and correlation of noise in each of the receivers are described by a noise covariance matrix $\mathbf{\Lambda_z}$, which reads as the covariance $\mathbf{\Lambda}$ of the measured data $\mathbf{z}$, so the noise follows $\mathscr{CN}(\boldsymbol{\mu_z}, \mathbf{\Lambda_z})$ where $\boldsymbol{\mu_z} = 0$. Image reconstruction of each of the DWI volumes $n \in \mathscr{N} = \{1, ..., N\}$ can be performed independently by linear inversion —a.k.a. sensitivity encoding (SENSE) method (Pruessmann et al., 1999)—:

$$\mathbf{y}_n^{\text{SENSE}} = \left(\mathbf{E}_n^H \mathbf{\Lambda_z}^{-1} \mathbf{E}_n\right)^{-1} \mathbf{E}_n^H \mathbf{\Lambda_z}^{-1} \mathbf{z}_n \qquad (1)$$

with $\mathbf{E}_n$ the encoding operator.

Assuming for now that the encoding is the same for all volumes, the basic encoding operator in parallel MRI can be expressed as $\mathbf{E} = \mathbf{FS}$, with $\mathbf{F}$ accounting for Fourier encoding (encompassing the applied spectral sampling structure) and $\mathbf{S}$ for sensitivity encoding. A whitening transformation $\mathbf{\Upsilon_z}$ of the measured receiver noise satisfying $\mathbf{\Upsilon_z}^H \mathbf{\Upsilon_z} = \mathbf{\Lambda_z}^{-1}$ is used in practice to transform the data $\bar{\mathbf{z}} = \mathbf{\Upsilon_z} \mathbf{z}$ and the encoding operator

$\overline{\mathbf{E}} = \mathbf{\Upsilon}_{\mathbf{z}}\mathbf{E} = \mathbf{F}\mathbf{\Upsilon}_{\mathbf{z}}\mathbf{S} = \mathbf{F}\overline{\mathbf{S}}$. This generates an equivalent inverse problem where the channel noise is decorrelated and standardized (Pruessmann et al., 2001), so we can write $\mathbf{\Lambda}_{\overline{\mathbf{z}}} = \mathscr{C}\mathscr{N}(\mathbf{0}_{KLN}, \mathbf{I}_{KLN})$, with $\mathbf{0}_{KLN}$ and $\mathbf{I}_{KLN}$ denoting respectively the null vector and identity matrix of size KLN, where K is the number of spectral samples and L is the number of coils.

Due to reconstruction linearity, $\mathbf{y}$ follows $\mathscr{C}\mathscr{N}(\boldsymbol{\mu}_{\mathbf{y}}, \mathbf{\Lambda}_{\mathbf{y}})$. The expected value of noise in the reconstructed data is simply

$$\boldsymbol{\mu}_{\mathbf{y}}^{\text{SENSE}} = \left(\overline{\mathbf{E}}^H\overline{\mathbf{E}}\right)^{-1}\overline{\mathbf{E}}^H\mathbf{0}_{KL} \otimes \mathbf{1}_N = \mathbf{0}_Q \otimes \mathbf{1}_N = \mathbf{0}_{QN}, \tag{2}$$

with $\otimes$ denoting the Kronecker product and $\mathbf{1}_N$ the all-ones vector of size N. On the other hand, the noise covariance is the same for all volumes:

$$\mathbf{\Lambda}_{\mathbf{y}}^{\text{SENSE}} = \left(\overline{\mathbf{E}}^H\overline{\mathbf{E}}\right)^{-1}\overline{\mathbf{E}}^H\mathbf{I}_{KL}\overline{\mathbf{E}}\left(\overline{\mathbf{E}}^H\overline{\mathbf{E}}\right)^{-1} \otimes \mathbf{I}_N = \left(\overline{\mathbf{E}}^H\overline{\mathbf{E}}\right)^{-1} \otimes \mathbf{I}_N = \mathbf{\Sigma}_Q \otimes \mathbf{I}_N, \tag{3}$$

with $\mathbf{\Sigma}_Q$ referring to a given covariance matrix of size Q. As for patch-based denoising, we are interested on the local patch marginalization of the noise covariance:

$$\mathbf{\Lambda}_{\mathbf{Y}^{[q]}} = \mathbf{P}^{[q]}\mathbf{\Lambda}_{\mathbf{y}}\mathbf{P}^{[q]H} = \mathbf{P}^{[q]}\mathbf{\Sigma}_Q\mathbf{P}^{[q]\mathbf{H}} \otimes \mathbf{I}_N = \mathbf{\Sigma}_M^{[q]} \otimes \mathbf{I}_N \tag{4}$$

Before building the random matrix recovery model, let's review the properties of noise in different reconstruction scenarios:

- **Additive noise.** This property will generally hold for complex reconstructed data, as the application of a linear reconstruction operator preserves noise additivity (2). However, this additive property is broken for magnitude-only data due to the non-linearity of the magnitude operation, particularly at low SNR. Using magnitude data, the noise standard deviation will depend on the signal level in agreement to a Rician distribution (Gudbjartsson and Patz, 1995), which prevents a straightforward formulation of the matrix denoising problem. In a somewhat naive manner, one could still apply the singular value shrinkage to magnitude data but, as we will illustrate in § 3.6, for low SNR the noise bias would blend into the signal components so, even if there is some noise reduction, the diffusion contrast appears hampered.

- **Full noise independence.** The simplest situation occurs when the Fourier encoding operator is given by the discrete Fourier transform (DFT), $\mathbf{F} = \mathcal{F}$, which corresponds to Cartesian sampling without acceleration and with matched acquired and reconstructed grids (identified hereafter as case FULL). Due to unitarity of the DFT operator, $\mathbf{F}^H = \mathcal{F}^{-1}$, the spatial noise covariance can be written as

$$\mathbf{\Lambda}_{\mathbf{y}}^{\text{FULL}} = \left(\overline{\mathbf{S}}^H\overline{\mathbf{S}}\right)^{-1} \otimes \mathbf{I}_N = \mathbf{\Sigma}_Q^{\text{FULL}} \otimes \mathbf{I}_N. \tag{5}$$

$\overline{\mathbf{S}}$ is a spatially diagonal operator, so it acts on each sampled voxel q separately. Then, it can be arranged for each voxel as a column vector of size $C \times 1$, $\overline{\mathbf{S}}_q = \overline{\mathbf{s}}_c$, and $\mathbf{\Sigma}_Q^{\text{FULL}}$ reduces to a diagonal matrix comprised of scalar variances:

$$\mathbf{\Sigma}_Q^{\text{FULL}} \equiv \mathbf{\Sigma}_1^{\text{FULL},\mathscr{C}} = \sigma_q^2 = \left(\overline{\mathbf{S}}_q^H\overline{\mathbf{S}}_q\right)^{-1}, \tag{6}$$

with σ_q^2 referring to the elements in the diagonal.

- **Local independence.** A common reconstruction setting is that of parallel imaging acceleration via homogeneous spectral under-sampling (case UNDER). In this case the Fourier encoding operator can be expressed as $\mathbf{F} = \mathcal{F}\mathbf{U}$, with $\mathbf{U}$ a spatial overlapping operator from the reconstruction to the acquisition field of view (FOV). Accordingly, the noise covariance is given by:

$$\mathbf{\Lambda}_{\mathbf{y}}^{\text{UNDER}} = \left(\overline{\mathbf{S}}^H\mathbf{U}^H\mathbf{U}\overline{\mathbf{S}}\right)^{-1} \otimes \mathbf{I}_N = \mathbf{\Sigma}_Q^{\text{UNDER}} \otimes \mathbf{I}_N. \tag{7}$$

The spatial noise correlation can be represented by a sparse matrix

with non-zero entries only along the main diagonal and the $U - 1$ element pairs corresponding to U aliased voxels in the reduced acquisition FOV. Considering a given aliased dimension i, these correlations will only affect those voxels separated by uK_i, with $u \in \mathscr{U} = \{0, ..., U - 1\}$ and K_i the number of acquired samples along dimension i. One can define a sensitivity matrix[1] $\overline{\mathbf{S}}_{\mathbf{q}_{\mathscr{U}}} = \overline{\mathbf{s}}_{c,u}$ for the overlapped voxels $\mathbf{q}_{\mathscr{U}} = \{q_0, ... q_{U-1}\}$ so that noise covariances are given by

$$\mathbf{\Sigma}_Q^{\text{UNDER}} \equiv \mathbf{\Sigma}_U^{\text{UNDER},\mathscr{C}/\mathscr{U}} = \sigma_{q_{u_1}, q_{u_2}}^2 = \left(\overline{\mathbf{S}}_{\mathbf{q}_{\mathscr{U}}}^H\overline{\mathbf{S}}_{\mathbf{q}_{\mathscr{U}}}\right)^{-1}, \tag{8}$$

with $\mathscr{C}/\mathscr{U}$ the quotient of $\mathscr{C}$ by $\mathscr{U}$. The local covariance matrix in (4) becomes diagonal due to local independence, and we can write $\mathbf{\Sigma}_1^{\text{UNDER},\mathscr{C}} = \sigma_{q_{\mathscr{U}},q_{\mathscr{U}}}^2$, computed using the set of overlapped voxels that includes the voxel q. Normalization by this matrix has been previously proposed as the reconstruction in SNR units (Kellman and McVeigh, 2005) and, for accelerated acquisitions, the ratio between $\sigma_{q_{\mathscr{U}},q_{\mathscr{U}}}$ and σ_q corresponds to the g-factor times the square root of the acceleration ratio at voxel q (Pruessmann et al., 1999).

- **Spatial correlations.** More general settings arise for more complex reconstruction techniques or when the data is acquired in a non uniform manner. Typical examples include when the Fourier operator involves regridding, or Gibbs ringing filtering and/or zero-filling interpolation are applied to the reconstructed data. These situations will generally break the noise independence. The specific noise properties will depend on the reconstruction operator. However, it is generally possible to completely characterize the noise propagation through a linear reconstruction method by making use of a given noise covariance matrix over the local patches, as defined in (4), which can be obtained either analytically or by Monte Carlo simulation. Here, as an example of interest in our data, we focus on partial Fourier (PF) or half scan acquisitions (case HALF) where the reconstruction model in (1) is extended by adding a spectral filter $\mathbf{G}$, for instance the ramp filter used for homodyne detection in (Noll et al., 1991) or simply a zero-filling filter. While standard PF reconstruction involves a separate phase estimation step, this can be integrated in the signal retrieval formulation within our setting, so we only need to include the k-space weighting required to retrieve the target signal resolution. The noise covariance becomes:

$$\mathbf{\Lambda}_{\mathbf{y}}^{\text{HALF}} = \mathbf{G}\left(\overline{\mathbf{S}}^H\mathbf{U}^H\mathbf{U}\overline{\mathbf{S}}\right)^{-1}\mathbf{G}^H \otimes \mathbf{I}_N = \mathbf{\Sigma}_Q^{\text{HALF}} \otimes \mathbf{I}_N. \tag{9}$$

- **Temporal heteroscedasticity.** This can be the case when the encoding structure is non stationary, typically from an acquisition technique where more evenly distributed net information is pursued by alternating the encodings for different excitations. Here we focus on the method we have used in our neonatal cohort (Hutter et al., 2018), where the phase encoding (PE) direction is interleaved volumewise among the four possible Cartesian in-plane directions for increased tolerance to main field inhomogeneity induced distortions (Bhushan et al., 2014). We can write the reconstruction for this case (case INTER) using the most general noise model of those previously described, HALF, as

$$\mathbf{y}_n^{\text{INTER}} = \mathbf{G}_{a(n)}\left(\overline{\mathbf{S}}_{a(n)}^H\mathbf{U}_{a(n)}^H\mathbf{U}_{a(n)}\overline{\mathbf{S}}_{a(n)}\right)^{-1}\overline{\mathbf{S}}_{a(n)}^H\mathbf{U}_{a(n)}^H\mathbf{F}^H\overline{\mathbf{z}}_n, \tag{10}$$

where $a(n)$ indexes the encoding used for volume n, with $a \in \mathscr{A} = \{1, ...,$

---

[1] This matrix is assumed to encompass the normalization from the acceleration ratio between accelerated and non-accelerated acquisitions, which is defined as the number of samples of a full acquisition divided by the number of samples of the accelerated scan.
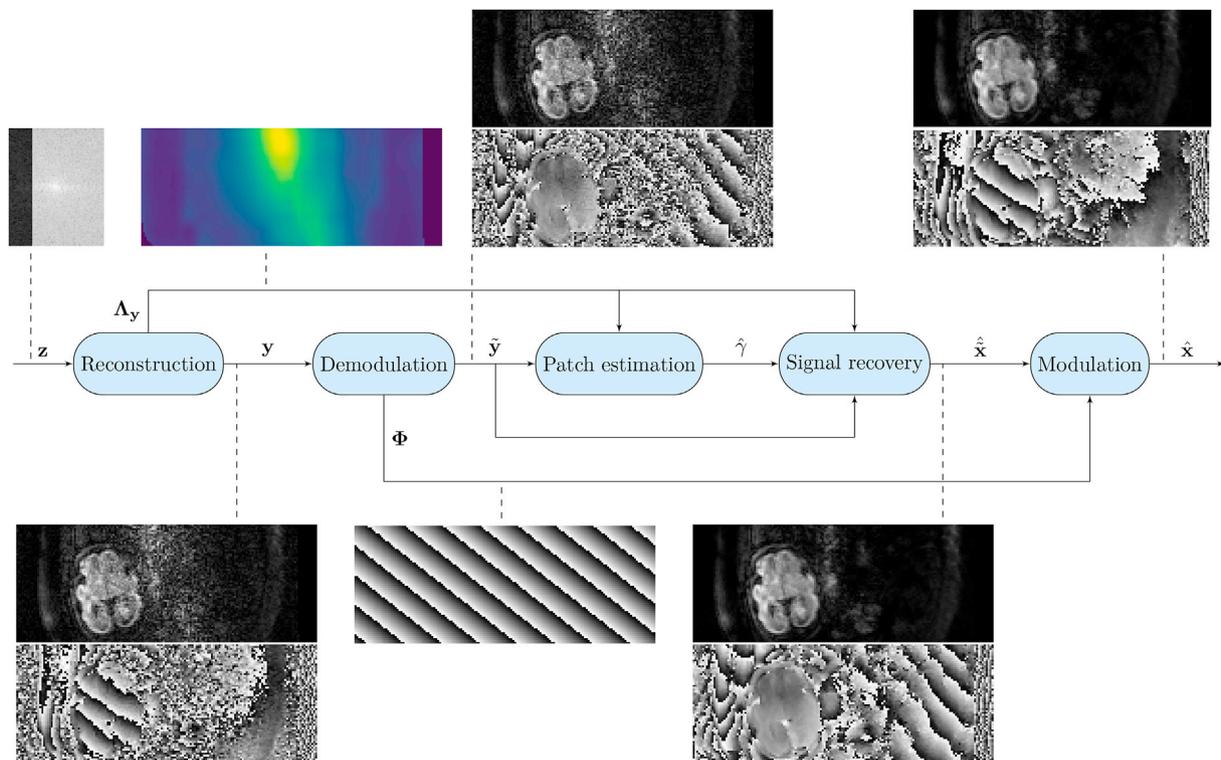
Fig. 1. Block diagram of the DWI denoising algorithm exemplified with a fetal dataset. Note PF data **z** has been acquired which produces a non-symmetric spectrum in the PE (horizontal) dimension. After the inversion of the DFT and spatial unfolding during the reconstruction, we have access to the magnitude and phase of **y** and the noise covariance $\Lambda_\mathbf{y}$, here illustrated by the spatial noise amplification levels. Later on, a linear phase $\mathbf{\Phi}$ is estimated and removed from the data, so we obtain $\tilde{\mathbf{y}}$. Optimal patch size $\hat{\gamma}$ is estimated to perform signal prediction, $\hat{\tilde{\mathbf{x}}}$, and phase demodulation is reversed to provide the final estimate $\hat{\mathbf{x}}$.

$A\}$. In this situation we can arrange the covariance between the matrix entries as a fourth order tensor:

$$\Lambda_\mathbf{y}^{\text{INTER}} = \sum_{a \in \mathscr{A}} \Sigma_Q^{\text{HALF},a} \otimes \Sigma_N^a, \tag{11}$$

where $\Sigma_Q^{\text{HALF},a}$ represents the spatial covariance matrix for the encoding a and Σ_N^a is an indicator diagonal matrix with ones in those entries that correspond to the volumes acquired with encoding a.

2.3. Denoising algorithm

Fig. 1 sketches the steps of our denoising procedure. Here is a detailed description:

- **Reconstruction.** We use the SENSE reconstruction framework in (1). Our datasets contain both in-plane and simultaneous multi-slice (SMS) accelerated echo planar imaging acquisitions, so the reconstruction method follows the extended SENSE technique in (Zhu et al., 2016). Noise correlation introduced by the readout gradient ramps is considered negligible. Sensitivity estimates from a conventional reference scan are refined with the information from non-SMS reference acquisitions with matched readouts (Hennel et al., 2016) to promote matched coil map and image distortions, which explains why the sensitivity matrices $\overline{\mathbf{S}}$ depend on the applied encoding in (10). Sensitivity estimation uses the variational formulation in (Allison et al., 2013).

- **Demodulation.** A model for phase corruption of the DWI signal in the presence of bulk motion was derived in (Anderson and Gore, 1994). Using this model, partial correction of the temporal phase variations can be achieved by estimating a linear phase $\mathbf{\Phi}$ for each of the reconstructed slices and subtracting it from the data, $\tilde{\mathbf{y}} = \mathbf{\Phi}^H \mathbf{y}$. By assuming stationarity and employing the Fourier shift theorem, the

linear phase ramp and offset are respectively estimated as the harmonic corresponding to the position of the maximum of the Fourier transform of the complex image and the phase at this maximum. This simple phase model is expected to be robust enough to be applied in low SNR regimes and to largely reduce the dynamic phase fluctuations, thus increasing the redundancy of the data and consequently reducing the rank of the matrices to be recovered. Phase correction of linear ramps in the PE direction may alter the temporal noise properties for PF, which could be incorporated into a model like (11), but for computational simplicity we assume that this effect is negligible.

- **Patch estimation.** The choice of the patch dimension M, or equivalently the matrix aspect ratio $\gamma = M/N$, may impact recovery performance. We propose to optimize the patch size by minimizing the relative asymptotic mean squared error (RAMSE) of matrix estimation over a set of candidate $\gamma_{(j)} : 0 < \gamma_{(j)} < \infty, j \in \mathscr{J} = \{1, ..., J\}$. The procedure is analogous to matrix recovery, so both will be described in § 2.4.

- **Signal recovery.** The optimal aspect ratio $\hat{\gamma}$ is used to build the patches for matrix recovery, which allows a denoised estimate $\hat{\tilde{\mathbf{x}}}$ to be obtained from the observed data $\tilde{\mathbf{y}}$.

- **Modulation.** The phase subtraction performed in the demodulation step can be reversed to obtain an estimate of the complex DWI signal, $\hat{\mathbf{x}} = \mathbf{\Phi}\hat{\tilde{\mathbf{x}}}$.

2.4. Matrix recovery

In this Section we describe both the procedure for matrix-based signal recovery and the related patch size estimation technique. The generic recovery pipeline is sketched in Fig. 2. Its inputs are the reconstructed and demodulated data from one of the reconstruction techniques described in § 2.2, $\tilde{\mathbf{y}}$, the noise covariance induced by the reconstruction, $\Lambda_\mathbf{y}$, and the optimal aspect ratio provided by the patch size estimation
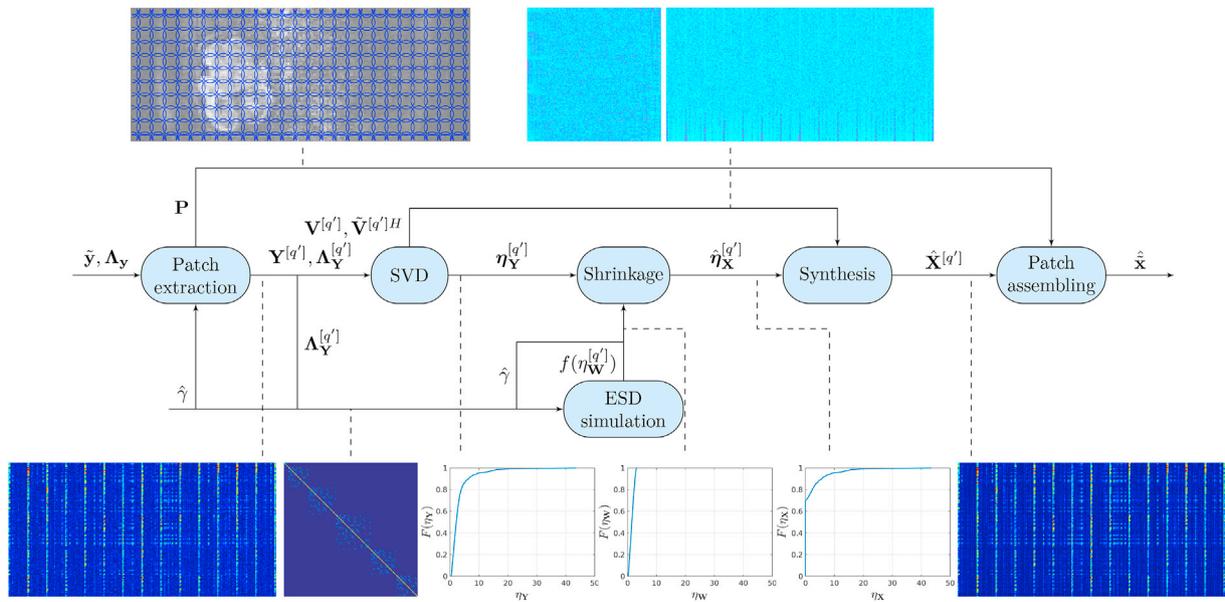
Fig. 2. Block diagram of the signal recovery procedure. Local information within ball-shaped sliding patches is extracted by **P**. Random signal plus noise matrices **Y** and covariance tensors $\mathbf{\Lambda_Y}$ are constructed for each of the patches $[q']$. SVD of these matrices is performed with singular vectors in **V** and $\tilde{\mathbf{V}}^H$ and singular values in $\boldsymbol{\eta_Y}$. In parallel, noise properties are modelled by a spectral distribution $f(\boldsymbol{\eta_W})$, which allows optimally shrunk singular values, $\hat{\boldsymbol{\eta}}_{\mathbf{X}}$, to be computed. Matrix estimates $\hat{\mathbf{X}}$ are synthesized using the shrunk singular values and overlapped patch information is assembled onto the signal estimates $\hat{\hat{\mathbf{x}}}$.

procedure described at the end of this Section, $\hat{\gamma}$. Here we summarize its blocks:

- **Patch extraction.** Patch extraction uses the procedure described in § 2.1. To accelerate computation, the number of patches to cover the FOV is lower than the total number of voxels, $Q' \leq Q$. In our default implementation, patch sliding is performed with a subsampling factor $T = 2$ in all directions, so $Q' \simeq Q/T^3$, with negligible impact in quality. This step outputs a patch extraction operator **P**, used later for patch assembling, the matrices to be recovered, with local information for a given patch arranged along the rows and corresponding diffusion measures along the columns, $\mathbf{Y}^{[q]}$, and a set of local noise covariance matrices, $\mathbf{\Lambda_Y}^{[q]}$.

- **SVD.** SVDs are computed for each patch matrix $[q']$, $\mathbf{Y}^{[q]} = \mathbf{V}^{[q]} \mathbf{\Theta}^{[q]} \tilde{\mathbf{V}}^{[q]H}$, with $\mathbf{V}^{[q]}$ and $\tilde{\mathbf{V}}^{[q]H}$ containing respectively the left and right singular vectors and $\mathbf{\Theta}^{[q]}$ a diagonal matrix with the singular values $\boldsymbol{\eta}_{\mathbf{Y}}^{[q]}$ in its diagonal.

- **Shrinkage.** The matrix recovery problem via SVD is grounded on the related spiked covariance model (Johnstone, 2001). This model was proposed to study the distribution of eigenvalues of sample covariance matrices[2] $N^{-1}\mathbf{YY}^H$ drawn from population covariances whose eigenvalues $\boldsymbol{\eta}_{\mathbf{X}}^2$ are all equal to 1 rather than for a small subset of them $\boldsymbol{\eta}_{\mathbf{X}r}^2$, $1 \leq r \leq R$, $R \ll M$, that satisfy $\boldsymbol{\eta}_{\mathbf{X}1}^2 \geq \dots \geq \boldsymbol{\eta}_{\mathbf{X}R}^2 > 1$ and are independent of the number of variables M and observations N. Then, defining an asymptotically constant ratio of variables and observations $\gamma = M/N$, it has been shown (Baik and Silverstein, 2006) that, whenever $\boldsymbol{\eta}_{\mathbf{X}r}^2 > 1 + \sqrt{\gamma}$, the spiky eigenvalues of the sample covariance matrix $\boldsymbol{\eta}_{\mathbf{Y}r}^2$ tend to a deterministic limit when $N \to \infty$ given by $\boldsymbol{\eta}_{\mathbf{Y}r}^2 = \boldsymbol{\eta}_{\mathbf{X}r}^2(1 + \gamma/(\boldsymbol{\eta}_{\mathbf{X}r}^2 - 1))$; i.e., there is a positive bias in the sample spiky eigenvalues as compared to the eigenvalues of the population $\boldsymbol{\eta}_{\mathbf{X}r}^2$. Building on this model, a set of optimal shrinkage criteria $\hat{\boldsymbol{\eta}}_{\mathbf{X}r}^2 = h(\boldsymbol{\eta}_{\mathbf{Y}r}^2)$ for different loss functions $L(\hat{\mathbf{X}}, \mathbf{X})$ have been derived in

(Donoho et al., 2018). In this paper we focus on the Frobenius loss criterion

$$L_{\text{FROB}}(\hat{\mathbf{X}}, \mathbf{X}) = ||\hat{\mathbf{X}} - \mathbf{X}||_2^2 \tag{12}$$

but keeping in mind that similar ways of proceeding are possible for other losses.

The condition of almost all population eigenvalues equal to 1 in (Donoho et al., 2018) makes this covariance estimation problem analogous to the matrix denoising problem in the presence of additive independent and identically distributed (i.i.d.) Gaussian noise, for which optimal shrinkers for different cost functions have been derived in (Gavish and Donoho, 2017). However, the description of noise propagation in the reconstruction in § 2.2 shows that the i.i.d. condition is generally not valid in DWI reconstructed data. (Benaych-Georges and Nadakuditi, 2012), using a more general model than (Gavish and Donoho, 2017), studied the asymptotic limits of the singular values of additive perturbations of zero-mean Gaussian not necessarily i.i.d. noise-only matrices **W**. They showed that, assuming a perturbation signal matrix drawn from a bi-unitarily invariant distribution, these limits only depend on integral transforms of the empirical spectral distribution (ESD) of noise, that is to say, on the distribution of the singular values of matrices with noise-only entries. Nadakuditi (2014) proposed to use those limits to build optimal shrinkers for the observed singular values when using unitarily invariant costs such as (12).

Following (Benaych-Georges and Nadakuditi, 2012), we can define the D-transform as

$$D_f(x) = \left(\int_{-\infty}^{\infty} \frac{x}{x^2 - \eta_{\mathbf{W}}^2} f(\eta_{\mathbf{W}}) d\eta_{\mathbf{W}} \right) \left(\gamma \int_{-\infty}^{\infty} \frac{x}{x^2 - \eta_{\mathbf{W}}^2} f(\eta_{\mathbf{W}}) d\eta_{\mathbf{W}} + \frac{1-\gamma}{x} \right), x > \eta_{\mathbf{W}}^+.$$

$$\tag{13}$$

with the ESD probability density $f(\eta_{\mathbf{W}})$ —or alternatively the ESD cumulative density $F(\eta_{\mathbf{W}})$— depending on the noise covariance matrix and the aspect ratio as described in the next item, and $\eta_{\mathbf{W}}^+$ referring to the upper bound of the support of $f(\eta_{\mathbf{W}})$. Then, it turns that from the set of all

---

[2] Without lack of generality we present the model for $M/N = \gamma < 1$.

possible shrinkers, the one that asymptotically minimizes the expected Frobenius loss in (12) can be obtained in terms of the ratio of the D-transform of the observed singular values above the detection threshold and its derivative (Nadakuditi, 2014):

$$
\widehat{\eta}_{\mathbf{X}r} = \begin{cases} -2 \dfrac{D_f(\eta_{\mathbf{Y}r})}{D'_f(\eta_{\mathbf{Y}r})} & \text{if } \eta_{\mathbf{Y}r} > \eta_{\mathbf{W}}^+ \\ 0 & \text{otherwise,} \end{cases} \tag{14}
$$

Note that defining the set of non-fully suppressed components $\mathscr{R} = \{r : \eta_{\mathbf{Y}r} > \eta_{\mathbf{W}}^+\}$, a rank estimate is obtained as $\widehat{R} = |\mathscr{R}|$, with $|\cdot|$ denoting the set cardinality. Consequently, in contrast to other DWI denoising techniques (Pai et al., 2011; Manjón et al., 2013; Veraart et al., 2016b), our proposal does not impose either truncation or full preservation of the SVD components, as truncation thresholds, even when optimally chosen, have been shown to compare unfavourably with optimal shrinkage (Gavish and Donoho, 2014, 2017). The benefit of shrinking with respect to truncating is relevant in situations where a given component may include comparable contributions of signal and noise, which we have seen to generally occur in our datasets.

- **ESD modelling.** The Marčenko-Pastur theorem (Marčenko and Pastur, 1967; Silverstein, 1995) states that the ESD of the population covariance matrix $\mathbf{\Sigma}_M$ converges almost surely to a bounded distribution that can be described by a fixed point equation. Numerical tools (Dobriban, 2015; Ledoit and Wolf, 2017) have been developed to compute the ESD for noise covariance matrices of the form $\mathbf{\Lambda}_{\mathbf{Y}} = \mathbf{\Sigma}_M \otimes \mathbf{I}_N$, which applies to the stationary encoding models in (6), (8) and (9). Although analogous fixed point equations (Wagner et al., 2012) and eigenvalue confinement guarantees (Kammoun and Alouini, 2016) are available for more general noise models such as those of non stationary encodings in (11), we do not know of any numerical tool developed for general random interactions between the matrix entries. Thereby, here we adopt a more flexible approach that approximates the asymptotic spectrum of noise by Monte Carlo simulations (Jing et al., 2010). Namely, we draw Q' matrices $\mathbf{W}$ of size $B(M \times N)$, with B a positive integer factor to guarantee that the spectrum of these matrices is a good enough approximation to their asymptotic limit. If these matrices are synthesized so that they follow the noise distribution in the data, $\mathscr{CN}(0, \mathbf{\Lambda}_{\mathbf{Y}^{[q]}})$, we can compute their decreasingly sorted singular values $\widehat{\boldsymbol{\eta}}_{\mathbf{W}}^{[q]}$ and make the following approximation for each patch $[q']$:

$$
f(\eta_{\mathbf{W}}) \simeq \frac{1}{BM} \sum_{m=1}^{BM} \delta(\eta_{\mathbf{W}} - \widehat{\eta}_{\mathbf{W}m}), \tag{15}
$$

with $\eta_{\mathbf{W}}^+ \simeq \widehat{\eta}_{\mathbf{W}1}$ and δ the Dirac delta. Note that, as a result of our investigations to directly compute the asymptotic noise spectrum, we have also developed some techniques for solving the generalized Marčenko-Pastur equation in the setting of (11). Although due to computational inefficiency these developments have not been adopted for our problem, we have collected them in a companion note (Cordero-Grande, 2019).

- **Synthesis.** Estimated signal matrices are reconstructed for each patch using the shrunk singular values, $\widehat{\mathbf{X}}^{[q]} = \mathbf{V}^{[q]} \widehat{\mathbf{\Theta}}^{[q]} \tilde{\mathbf{V}}^{[q]H}$, with $\widehat{\mathbf{\Theta}}^{[q]}$ a diagonal matrix built from the singular values $\widehat{\boldsymbol{\eta}}_{\mathbf{X}}^{[q]}$ obtained by (14).
- **Patch assembling.** The estimated DWI volumes are obtained as

$$
\widehat{\mathbf{x}} = \widehat{\tilde{x}}_{q_2,n} = \sum_{q_1 \in \mathscr{C}'} \omega_{q_2}^{[q_1]} \widehat{X}_{m \to q_2, n}^{[q_1]}, \tag{16}
$$

with $\omega_{q_2}^{[q_1]}$ satisfying $\sum_{q'_1 \in \mathscr{C}'} \omega_{q_2}^{[q'_1]} = 1$ used to weight the matrix estimate

from the patch centered at q_1 in the q_2 assembled estimate. This weighting can be chosen to be uniform, depending on the inverse of the variance of the estimates, as in (Dabov et al., 2007), which in our setting could be obtained from (17) below, or based on the distance between q_2 and q_1, for instance by means of a Gaussian window. Simulation tests in § 3.3 were run using inverse variance combination, which provided the best results. However, the differences with both uniform and Gaussian kernel combination were small, so in the in vivo tests in §§ 3.4, 3.5, and 3.6 we opted for a Gaussian window combination, which was judged potentially beneficial in the presence of motion.

The procedure for patch size estimation is sketched in Fig. 3. It uses the same inputs and analogous methods than the decomposition for matrix recovery, but the objective is to select the patch size that yields the best RAMSE. In what follows we describe the implementation details and the formulas for RAMSE estimation:

- **Patch extraction.** This step is analogous to that for actual denoising but a smaller number of patches Q'' is used to cover the FOV. Drawing a moderate number of patches will generally provide enough information for reliable global RAMSE estimates. Our in vivo tests have used a subsampling factor $T_\gamma = 6$ in each direction. Using this factor, the computation time of patch estimation has remained below the time for denoising.
- **SVD/ESD modelling.** These steps are the same as for actual denoising.
- **RAMSE estimation.** This step uses additional results provided in (Nadakuditi, 2014) to obtain estimates of the asymptotic mean squared error (AMSE) of matrix recovery based on integral transforms of the ESD evaluated at the observed singular values above the detection threshold. Namely, for the optimal shrinker in (14)

$$
\widehat{L}_{\mathrm{FROB}}^f(\widehat{\mathbf{X}}, \mathbf{X}) = \sum_{r:\eta_{\mathbf{Y}r} > \eta_W^+} \frac{1}{D_f(\eta_{\mathbf{Y}r})} - \widehat{\eta}_{\mathbf{X}r}^2. \tag{17}
$$

To determine the optimal aspect ratio, the errors are normalized as follows

$$
\widehat{\overline{L}} = \frac{\sum_{q'' \in \mathscr{C}''} \widehat{L}_{\mathrm{FROB}}^f(\widehat{\mathbf{X}}^{[q'']}, \mathbf{X}^{[q'']})}{\sum_{q'' \in \mathscr{C}''} \|\widehat{\mathbf{X}}^{[q'']}\|_2^2}. \tag{18}
$$

This feature is computed for a set of candidate $\gamma_{(l)}$ and the $\widehat{\gamma}$ offering the minimum RAMSE $\widehat{\overline{L}}$ is chosen for denoising. Importantly, these asymptotic error estimates can be used to assess the expected performance of different denoising alternatives in vivo without requiring a ground truth, which will be a key ingredient in § 3. Finally, note that the inverse of the RAMSE provides an estimate of the SNR after denoising.

3. Validation and results

In this Section we validate the different contributions of our framework for the brain DWI domains and specific sequences described in § 3.1 and using simulations. In § 3.2 we show the robustness of the noise measurement and propagation approach. Simulation-based experiments comparing our GSVS method with state of the art techniques are included in § 3.3. Improved recovery by using phase correction is tested in § 3.4. Effects and pertinence of applying our tools for generic noise models are studied in § 3.5. Finally, in § 3.6 we visually compare GSVS with the related MPPCA method in vivo.

3.1. Explored applications

A series of experiments have been conducted including three distinct brain DWI application domains: high *b*-value exploration in adult brain imaging, multi-shell high angular resolution neonatal examinations, and multi-shell fetal imaging. All the volumes in the experiments have been
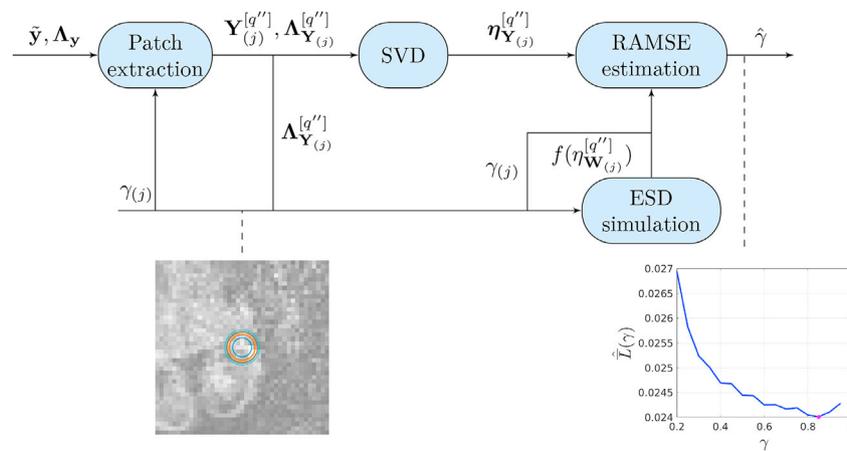
Fig. 3. Block diagram of the patch estimation procedure. Averaged RAMSE estimates $\widehat{L}$ over a subset of the patches used for denoising $\mathscr{C}''$ are computed for a set of candidate matrix aspect ratios $\gamma_{(j)}$ (which induce a set of patch radii). The candidate offering the minimum RAMSE $\widehat{\gamma}$ is used for denoising.

Table 1
Sequence parameters for each of the DWI application domains. acc refers to acceleration factors.

			Adult		
Resolution (mm^3)	FOV$_{rec}$ (mm^3)	PE acc	SMS acc	PE shift	b-values (ms /μm^2)
2.5 × 2.5 × 2.5	240 × 240 × 100	2	2	1/2	0/0.1/0.4/0.9/1.6/2.5 /3.6/4.9/6.4/8.1/10
Slice distance (mm)	PEs	T_R (s)	T_E (ms)	PF	# images/b-value
2.5	AP	6.9	99	1	15/16/19/24/31/40 /51/64/79/96/115
			Neonatal		
Resolution (mm^3)	FOV$_{rec}$ (mm^3)	PE acc	SMS acc	PE shift	b-values (ms /μm^2)
1.5 × 1.52 × 3	150 × 150 × 96	1.19	4	1/2	0.0/0.4/1.0/2.6
Slice distance (mm)	PEs	T_R (s)	T_E (ms)	PF	# images/b-value
1.5	LR/RL/AP/PA	3.8	90	0.86	20/64/88/128
			Fetal		
Resolution (mm^3)	FOV$_{rec}$ (mm^3)	PE acc	SMS acc	PE shift	b-values (ms /μm^2)
2.0 × 2.0 × 2.0	200 × 312 × 108	2	2	1/2	0.0/0.4/1.0
Slice distance (mm)	PEs	T_R (s)	T_E (ms)	PF	# images/b-value
2.0	AP	6.2	76/137	0.75	16/45/80

acquired using a 3 T Philips Achieva TX with $C = 32$-channel coils. Written informed consent for each participant was obtained prior to them being scanned. Neonatal and fetal study procedures were reviewed and approved by the Riverside Research Ethics Committee (14/LO/1169). DWI data was acquired as part of a broader examination aimed to study brain development within the developing Human Connectome Project (dHCP, 2019). Main sequence parameters are reported in Table 1 using a sequence design that allows for flexible sampling of the diffusion space, interleaved choice of the PE direction, free temporal arrangement of the sampling scheme, usage of SMS acceleration, and restart capacity (Hutter et al., 2018).

DWI recovery in these domains confronts different challenges. In the adult case, the high b-value DWI signal may fall below the Rician noise floor so the complex valued signal model seems advisable. In the neonatal case, we encounter a challenging acquisition protocol, where high spatial resolution DWI information has been acquired along four different PE directions interleaved in the diffusion space, using a SMS factor of 4 and with large distortions due to low in-plane acceleration. In addition, these datasets are frequently hampered by motion artifacts,

partly because the newborns have been scanned without sedation. Finally, in fetal examinations the targeted brain data is collected from locations that may be far away from the coils, with consequent low SNR levels. In addition, there is breathing motion and perhaps fetal motion throughout the acquisition, so that retrieval and representation of DWI information becomes extremely challenging. Two T_E values are reported for this dataset as it has been acquired using a dual hybrid spin echo and field echo sequence where the dead time after the application of the spin echo is used to collect some field echo information that is later used to correct for dynamic main magnetic field variations, which will be the subject of a future publication. Both echoes are acquired with the exact same encoding structure, so we can easily incorporate the multiple echo information to our denoising technique, as previously proposed in (Bydder and Du, 2006).

3.2. Propagation of noise measures

When the noise level is not available before denoising, different techniques have been proposed for its estimation. In (Veraart et al.,

Table 2
Considered noise estimation techniques.

Technique/Reference	Noise estimate
Spectral expectation (Veraart et al., 2016b)	$\tilde{\sigma}^2_{\text{EXP1}}(R) = \dfrac{\sqrt{N}(\eta^2_{YR+1} - \eta^2_{YM})}{4\sqrt{M-R}}$
	$\tilde{\sigma}^2_{\text{EXP2}}(R) = \dfrac{\sqrt{N-R}(\eta^2_{YR+1} - \eta^2_{YM})}{4\sqrt{M-R}}$,
	$\widehat{R}_{\text{EXP}\{1,2\}} = \inf\left\{ R \,\middle|\, \sum_{r=R+1}^{M} \eta^2_{Yr} \geq (M-R)\tilde{\sigma}^2_{\text{EXP}\{1,2\}}(R) \right\}$ (19)
	$\hat{\sigma}^2_{\text{EXP}\{1,2\}} = \tilde{\sigma}^2\left(\widehat{R}_{\text{EXP}\{1,2\}}\right)$
Spectral median (Gavish and Donoho, 2014)	$\hat{\sigma}^2_{\text{MED}} = \dfrac{\eta^2_{YM/2}}{\theta^2_{M/2}}$ (20)

2016b) the underlying matrix rank $\widehat{R}$ and the noise level $\hat{\sigma}$ are simultaneously estimated using statistics on the expectation of the Marčenko-Pastur distribution for noise matrices with additive i.i.d. entries. This estimator is identified hereafter as EXP1. The procedure is summarized by (19) within Table 2; increasing signal ranks R are generated until a given criterion for the rank estimate depending on standardized and decreasingly sorted eigenvalues is satisfied, with the standardization performed using the estimate for σ^2 at the current signal rank, $\tilde{\sigma}^2(R)$. We have studied this estimator and observed a biased behaviour when applied to close to square matrices, so we propose a simple modification that we identify as EXP2 in (19). On the other hand, in (Gavish and Donoho, 2014) a simple estimator is proposed where the standard deviation of noise is computed as the ratio of the median of the observed eigenvalues, $\eta^2_{YM/2}$, and the median of the Marčenko-Pastur law, $\theta^2_{M/2}$, as shown in (20). This estimator will be identified as MED.

To show the correctness of noise propagation through the reconstruction as well as the limitations of these data-based noise estimation methods, we use the full Fourier adult dataset (PF factor of 1 in Table 1). In this case, noise is captured by (7). Strictly speaking, the noise level within a given patch is not unique as it changes according to the spatial variation of the receiver field and the g-factor profiles, with the latter potentially inducing abrupt changes in the noise properties of neighboring locations. To avoid this inherent limitation of data-based noise level estimation when operating in signal units, our pipelines are modified in this experiment so that the data is standardized by the application of a whitening transformation Υ_M that normalizes by the noise factors. After applying this transformation, optimal asymptotic loss guarantees

are pursued on the invariant Frobenius loss (Canu and Fourdrinier, 2017), $L_{\text{FROB},\Upsilon_M}(\widehat{\mathbf{X}}, \mathbf{X}) = ||\Upsilon_M(\widehat{\mathbf{X}} - \mathbf{X})||^2_2$. If our noise measurement and propagation model is correct, the estimated noise level after standardization should be $\hat{\sigma} = 1$ everywhere. We check whether this is the case by running the estimators in Table 2 for reconstructions of noise-only data with the noise injected in place of the original data after the channel noise standardization described in § 2.2. These results are compared with the estimations obtained when reconstructing the original signal plus noise data.

Fig. 4a compares the cumulative histograms of noise estimates at the different spatial locations. Results show that when estimating noise levels in the absence of additive signal perturbations (cases $\hat{\sigma}^{\mathbf{W}}$) the distribution of estimates takes the value 1 everywhere, with only minor numerical deviations. This shows both the consistency of the estimators in Table 2 and the correctness of our noise propagation method. However, in the presence of signal perturbations (cases $\hat{\sigma}^{\mathbf{Y}}$), the estimators in Table 2 start to deviate from their asymptotic behaviour, with positive biases for EXP2 and MED and potentially larger dispersions for EXP1, which shows the limitations of data-based noise estimation techniques. Fig. 4b and c show exemplary spatial maps with the estimated noise levels for the noise-only data and the signal plus noise data using the shape corrected version of (19), EXP2, which has been judged to provide the most benign dispersion and bias tradeoff according to Fig. 4a. On the other hand, Fig. 4d and e compare the estimated ranks when using EXP2 and our noise propagation and asymptotic random matrix rank estimation. The results are in agreement with the elementary interplay where noise overestimation of EXP2 (Fig. 4c) is accompanied by lower rank estimates (Fig. 4d).

3.3. Simulation-based validation

A set of synthetic experiments have been conducted to quantitatively assess the performance of our method. Namely, we have used the *phantomas* software (Caruyer, 2019) to synthesize the 550 diffusion gradients corresponding to the adult protocol in Table 1 using the default diffusion parameters. Realistic SNR levels have been simulated by median-based harmonization of the $b = 0$ intensity levels in the phantom to those of the adult scan and noise generation according to the noise covariance after reconstruction. As we have already mentioned, most of current advanced denoising techniques are proposed for magnitude-only data, which complicates comparisons with the complex-domain methodology proposed here. This difficulty has been circumvented by performing comparisons within an idealized scenario where we assume that real-only information with complete restoration of Gaussianity has been made available by a given bias correction technique. Two state of the art
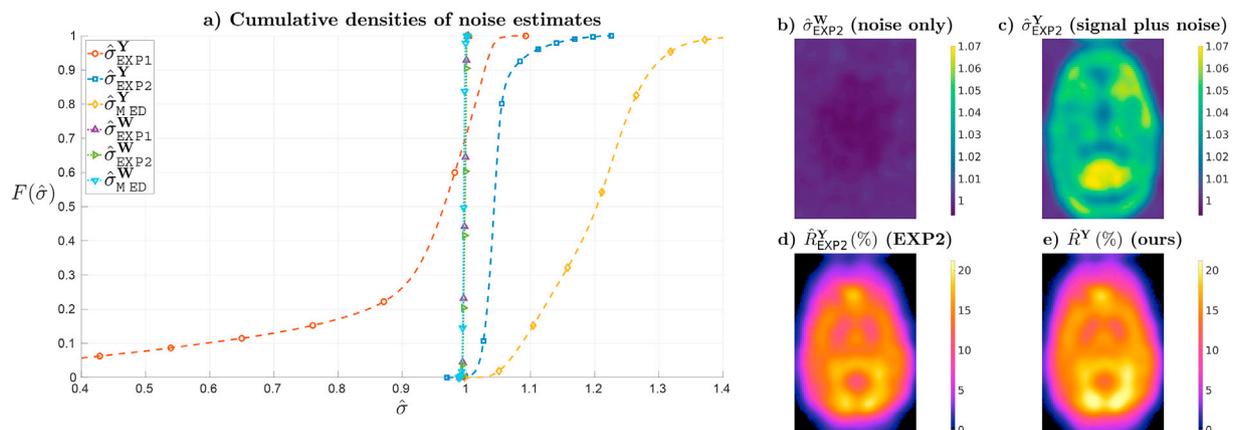

Fig. 4. a) Cumulative spatial distributions of noise estimates for different estimators without ($\hat{\sigma}^{\mathbf{W}}$) and with ($\hat{\sigma}^{\mathbf{Y}}$) signal perturbation. **b,c)** Spatial maps of noise for estimator EXP2 **b)** without and **c)** with signal perturbation. **d,e)** Rank estimates $\widehat{R}$, given as a percentage of the number of volumes N, for estimators **d)** EXP2 and **e)** ours. All results correspond to the RAMSE-optimal aspect ratio $\hat{\gamma} = 0.85$.

methods have been chosen for comparison, the related MPPCA in (Veraart et al., 2016b) and the non local spatial and angular matching (NLSAM) non-local means based approach in (St-Jean et al., 2016), to our knowledge the most recent non-local means approach for DWI with an available on-line implementation. NLSAM allows to input a noise standard deviation map, which enables more direct comparisons to our methodology. This is not the case for MPPCA, so we have resorted to the EXP2 noise estimation in (19). To discern the impact of noise estimation versus noise modeling, we have also applied our technique using the noise maps from the EXP2 patch estimates instead of the noise characterization derived from the reconstruction, which we term as the GSVS$\widehat{\sigma}$ alternative.

Fig. 5a and b display the peak SNR (PSNR) and structural similarity index (SSIM) perceptual metric (Wang et al., 2004) for each method as functions of the b-values in the simulated sequence. Results show that the GSVS approach ranks the best for all b-values both in terms of PSNR and SSIM. On average, the PSNR of the GSVS method improves by 7.82 dB when compared to MPPCA, with 3.96 dB basically attributable to noise modeling (difference with GSVS$\widehat{\sigma}$) and the remaining 3.86 dB due to the refinements in the statistical characterization of the problem. The difference rises to 16.32 dB and 20.81 dB when comparing to NLSAM and synthesized data respectively. In Fig. 5c we show a visual example for a slice with $b = 10 \text{ ms}/\mu\text{m}$ of an analogous experiment where the SNR of the original data is attenuated by 10 dB, as this SNR regime allows for easier perception of the differences between methods. Despite the original simulated data being severely degraded by noise, the underlying structures are visible after applying any of the denoising strategies but stronger noise suppression and fewer artifacts are noticeable when using the GSVS or GSVSσ approaches. Total computing time of our method has been of 194 s, with 34 s for patch size estimation and 160 s for actual denoising, which compares favourably with the 6836 s (8-core multiprocessing) employed by NLSAM and is similar to the MPPCA time. In Fig. 5d we look at the assumptions of the patch size estimation procedure as well as those of the methodology for in vivo comparisons. Using a subsampling factor of $T = 2$ as a baseline, we compare the global PSNR

$L_\mathbf{x}$ of GSVS (red solid) and MPPCA (green solid), the global PSNR of GSVS without patch overlap $L_\mathbf{X}$ (blue dashed) and the predicted global PSNR of GSVS $\widehat{L}_\mathbf{X}$ both for $T = 2$ (dotted yellow) and $T = 4$ (dotted magenta) obtained by the AMSE in (17). We observe that the profiles of the curves are similar for the range of compared aspect ratios γ with the GSVS outperforming MPPCA across the whole range. PSNR improvement due to patch assembling is also noticeable for all compared γ. The PSNR predictions generally provide a good approximation of the observed errors before patch assembling with estimates being practically equivalent for $T = 2$ and $T = 4$ as these are aggregates averaged across a large number of patches. However, there is a certain overestimation of errors for large aspect ratios. Closer inspection has revealed that this is due to relatively larger impact of deviations from the asymptotic assumptions for large γ. This limitation, probably related to the number of variables M being bigger than the number of observations N has motivated constraining the patch size estimation to $\gamma < 1$ for in vivo applications. Finally, the errors are relatively stable around the optimal γ so approximate techniques for patch size estimation seem to generally suffice in practice.

3.4. Effect of phase correction

In Fig. 6a we show the spatial distributions of the improvement ratio ρ in the RAMSE when using the full method (PC) versus removing the phase correction in § 2.3 (NPC), which is computed as $\rho = -10\log_{10}\widehat{\overline{L}}_{\text{PC}}/\widehat{\overline{L}}_{\text{NPC}}$. Plots are generated using brain-only information and they include the adult case and exemplary neonatal and fetal studies drawn at random. A reduction of the RAMSE is observed in all tested cases when using phase correction, with improvements ranging from around 0 to 1.5 dB depending on the location. In Fig. 6b–d we illustrate with comparisons of original data, results of denoising without phase correction and results of denoising with phase correction in an exemplary subject, slice and orientation within the highest b-value shell for the neonatal case. Despite the estimator operates on complex data, for the
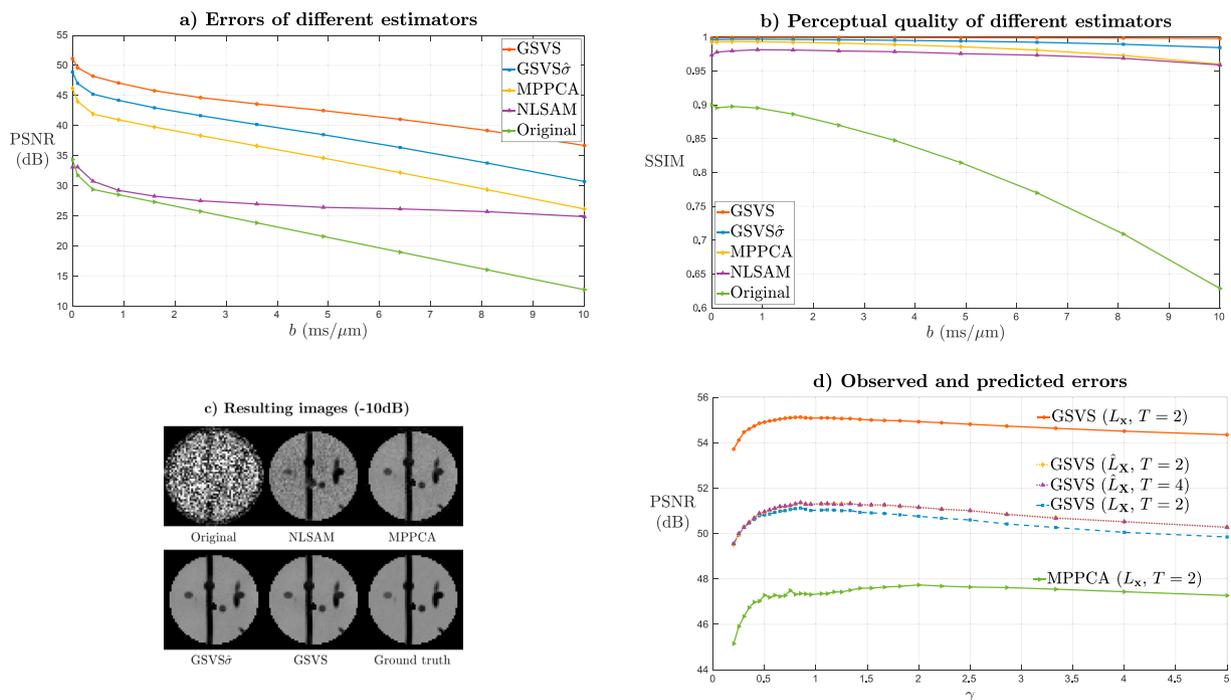

Fig. 5. Comparison of different denoising approaches on a synthetic phantom. **a)** PSNR and **b)** SSIM for the compared approaches and the original data at a baseline SNR that matches a real acquisition. **c)** Exemplary images with 10 dB SNR attenuation with respect to baseline when synthesizing the data. **d)** Measured PSNR in the resulting image (solid), together with measured (dashed) and predicted (dotted) PSNR in the matrices for different methods and patch subsampling factors.
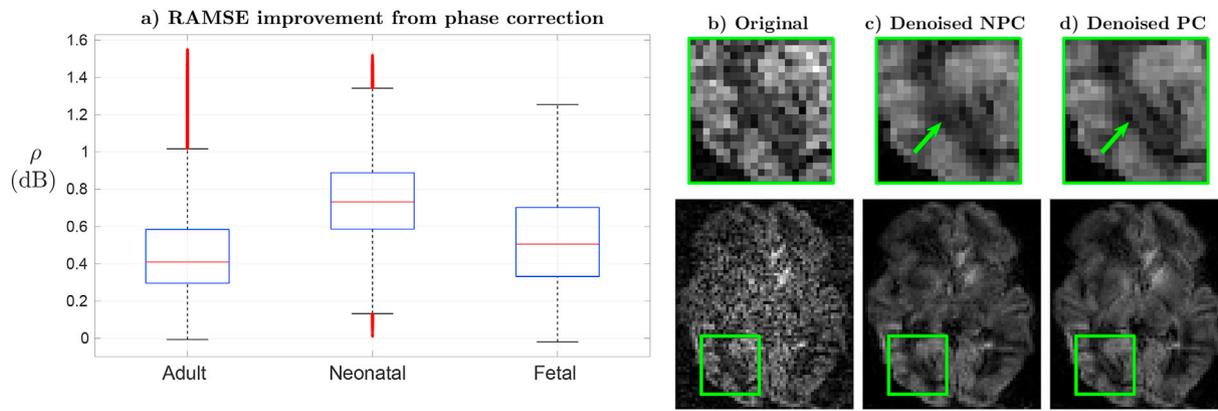
Fig. 6. a) RAMSE improvement ratio ρ of denoising with phase correction (PC) versus denoising without phase correction (NPC). **b-d)** NPC and PC denoising results for an exemplary slice at $b = 2.6\text{ms}/\mu\text{m}^2$ in the neonatal case: **b)** original data, **c)** NPC denoising, **d)** PC denoising. Same estimated aspect ratios $\hat{\gamma}$ have been obtained for NPC and PC denoising, respectively of 0.85, 0.8 and 0.9 for adult, neonatal and fetal data. Mean inverse RAMSEs with phase correction have been 20.52, 20.17 and 12.32 dB respectively for the adult, neonatal and fetal cases.

sake of economy we only provide magnitude displays of the results. Emergence of plausible anatomical features with better preserved resolution can be observed when using phase correction, for instance in the area of the posterior thalamic radiation indicated by the arrow.

3.5. Denoising with a generic noise model

The flexibility of the denoising scheme is illustrated by its application under different reconstruction operators and associated noise models. In a first experiment we use the fetal data to compare three approaches for PF reconstruction. In the first approach, hereafter referred to as regridded, the reconstructed PF data is regridded so that only the sampled area of the spectrum is back-transformed and consequently the data is represented in a coarser grid in the PE direction but with preserved local noise independence. In this case, the noise model follows (7) and the spatial patches are drawn elliptically rather than spherically. After denoising, the data is zero-filled and ramp-filtered as in (Noll et al., 1991) to retrieve the desired grid and resolution. As for the second approach, referred to as zero-filled, **G** in (9) takes the role of a zero filling filter, $\mathbf{G}_{\text{ZF}}$, with a ramp filter being applied after denoising. The third approach, referred to as ramped, applies the zero filling and ramp filters before denoising, so $\mathbf{G} = \mathbf{G}_{\text{RA}}$. Fig. 7a and b show, for the spin echo data at the highest b-value shell, the power spectral density (PSD) of the recovered

magnitude signal in both the readout and PE directions for the three denoising approaches and the original data. The PSD along a particular dimension is estimated by averaging along the remaining dimensions and the diffusion orientations. The PSDs generally follow a plausible linear decay after denoising. Overall, the ramped approach provides stronger noise suppression than the other two in the readout direction (as given by more pronounced PSD decay ratios) and similar suppression than the zero-filled approach in the PE direction, with weaker suppression and implausible shape for the regridded PE PSD at high frequencies. These results are in agreement with the implicit SNR-resolution tradeoffs of the different models used for denoising, with lower SNR of ramped when compared to zero-filled (Noll et al., 1991) and larger targeted resolution of zero-filled versus regridded. The low rank prediction adapts to the different SNR-resolution tradeoffs as reflected by the differences in resulting PSD profiles. The denoised images are compared in Fig. 7c–f, showing the original data (Fig. 7c) and the data denoised with the considered PF approaches. Regridded denoising (Fig. 7d) produces some spurious oscillations along the PE (horizontal) direction (see the area enclosed by the ellipse), which agrees with the shape of the PE PSD response at high frequencies. As for the zero-filled (Fig. 7e) and ramped (Fig. 7f) results, differences are more subtle but stronger background noise suppression can be appraised in the ramped case (see the area enclosed by the circle), which seems also in agreement with the slightly
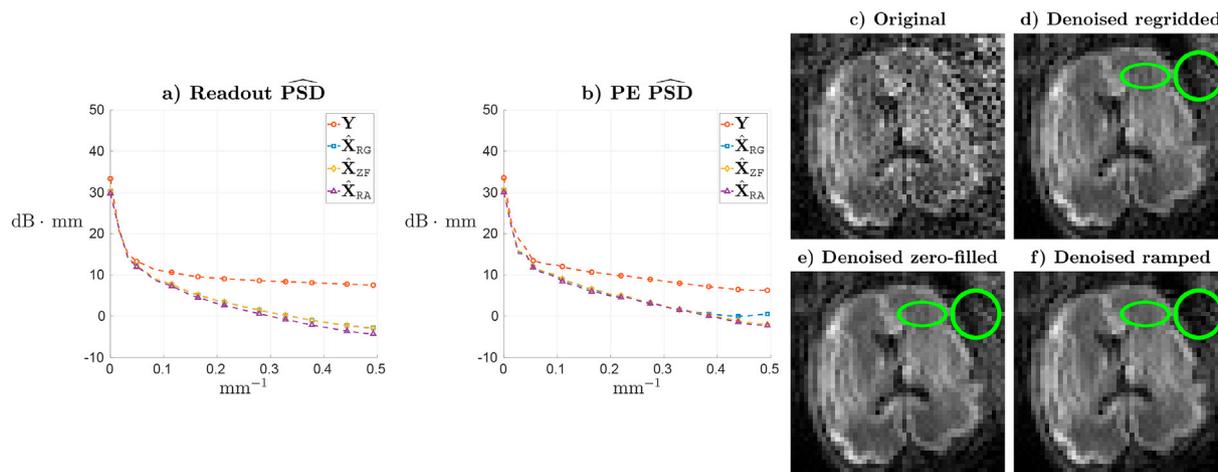

Fig. 7. a,b) PSD for non-denoised data, $\widehat{\text{PSD}}^{\mathbf{Y}}$, and the different denoising alternatives, $\widehat{\text{PSD}}^{\mathbf{X}_{\text{RG}}}$ (regridded), $\widehat{\text{PSD}}^{\mathbf{X}_{\text{ZF}}}$ (zero-filled) and $\widehat{\text{PSD}}^{\mathbf{X}_{\text{RA}}}$ (ramped), along the **a)** readout and **b)** PE directions. **c-f)** Resulting images **c)** before denoising, and when denoising using the **d)** regridded ($\hat{\gamma} = 0.65$), **e)** zero-filled ($\hat{\gamma} = 0.85$) and **f)** ramped ($\hat{\gamma} = 0.9$) approaches.
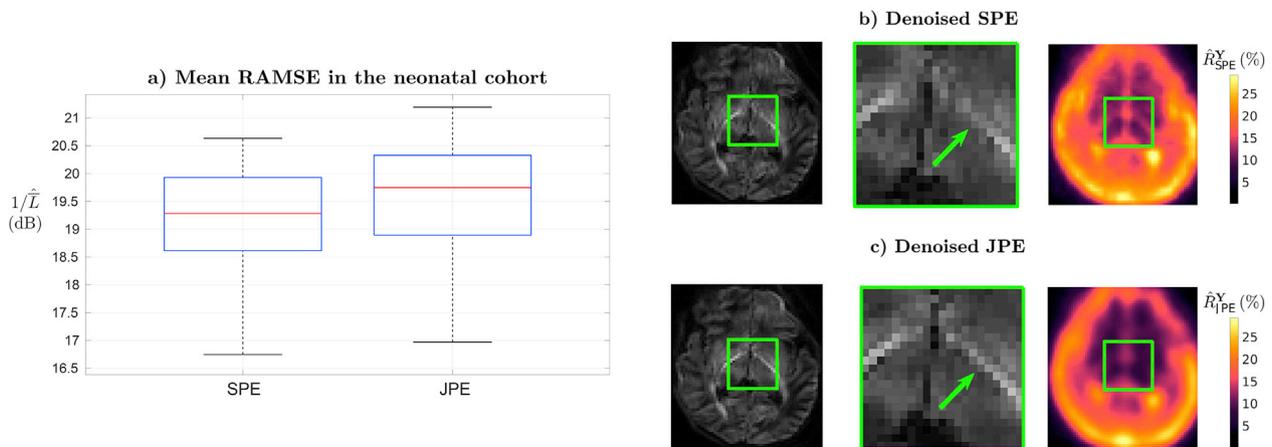
Fig. 8. a) Distributions of the estimated RAMSE for denoising using the volumes of each PE separately (SPE) and all volumes together (JPE). **b,c)** From left to right, denoising for **b)** SPE ($\hat{\gamma} = 0.95$ for all PEs) and **c)** JPE ($\hat{\gamma} = 0.8$) for an exemplary slice at $b = 2.6 \text{ms}/\mu\text{m}^2$; zoomed results; rank estimates $\hat{R}$, given as a percentage of the number of volumes N.

stronger decay of the readout PSD for this approach. These experiments show that our denoising method provides flexible ways of balancing the SNR-resolution tradeoff in PF reconstructions considering, for instance, the noise amplification introduced when ramp filtering in homodyne reconstructions. They also show that while it is possible to reduce the problem to one with independent noise samples along the local patches by manipulating the reconstruction grid, this has an impact on the results at the prescribed grid and resolution.

In the neonatal case, four PE directions were interleaved throughout the acquisition, each of them generating its own spatial noise distribution. Two strategies to deal with these datasets are compared. First, the problem is reduced to four independent denoising subproblems that use the volumes corresponding to each PE separately (SPE method). Second, the noise model in (11) is considered so the different PEs are treated jointly (JPE method). The comparisons are performed at a group level using a sample of 87 subjects chosen uniformly at random from the set of complete scans within the neonatal cohort and using $\mathbf{G} = \mathbf{G}_{\text{RA}}$ for both strategies. The inverse RAMSE averaged on automatically extracted brain voxels is compared in Fig. 8a. A reduction of the average RAMSE $\hat{\bar{L}}$ has been observed when using the JPE approach with an average improvement ratio of $\rho = 0.39$ dB. These results give a negligible p-value when performing a paired right-tailed sign test against the null hypothesis that the median of $\hat{\bar{L}}_{\text{JPE}} - \hat{\bar{L}}_{\text{SPE}}$ is greater than zero or zero. Left column of Fig. 8b and c shows that the recovered brain signal appears similar at coarse scales. However, when zooming in, as in the central column, better delineated edges are observed when using the JPE model, as observed in the ascending tract feature highlighted by the arrow. The right column explains this effect by resorting to the rank estimates, with JPE estimated signal ranks generally lower and smoother than the sum of SPE estimates for each PE; more data samples have been made available by leveraging all PEs so that any residual redundancy in the information contained in the different PEs will increment the denoising potential.

3.6. Complex denoising for unbiased diffusion measures

This Section provides visual evidence of the benefits of operating with complex data for denoising and performs comparisons with the MPPCA method in (Veraart et al., 2016b), which uses the EXP2 noise estimator in (19). To separately assess the technical refinements introduced in matrix recovery, comparisons are also performed when adapting this method to operate on complex data. Due to complexity of neonatal and fetal data preprocessing for motion and distortion correction and the limited applicability of MPPCA under general noise models, we use diffusion

measures of non-preprocessed adult data. Similar properties of magnitude and complex data retrieval have been observed in neonatal and fetal examples. Fig. 9 shows the denoising results for the $b = 10 \text{ms}/\mu\text{m}^2$ shell using an exemplary orientation (top) together with the estimated spherical harmonic power obtained with MRTRIX3 (Tournier et al., 2019) at orders $l = 0$ (mid) and $l = 8$ (bottom). We illustrate the results on an inferior slice with particularly low SNR and affected by strong non-linear phase corruption. The denoising capabilities are evident when comparing with the original data in Fig. 9a. However, the retrieved signal differs noticeably for the tested approaches. Magnitude-only denoising in Fig. 9b, using MPPCA, and 9c, adapting our GSVS method (assuming the same noise model but in a real instead of a complex domain), provides biased estimates when compared to complex denoising in Fig. 9d, adapting MPPCA, and 9e, using GSVS, especially in free diffusion areas. The bias reduces the contrast between restricted and free diffusion, with hampered detectability of the optical nerve for $l = 0$ (highlighted by the arrows). Complex denoising also appears more effective in preserving the diffusion features for $l = 8$, for instance at the cerebellum (highlighted by the arrows). Differences between data-based noise estimation and hard thresholding (Fig. 9d) versus propagation of noise measures and optimal shrinkage (Fig. 9e) are more subtle. However, it appears that the high spatial frequency harmonics are more accurately recovered when using the methods here proposed, with more compactly structured features in 9e versus 9d (top row and $l = 8$). This is supported by the random matrix error estimates, which predict an average AMSE reduction of 1.84 dB (ranging from 1.10 to 5.12 dB at different locations within the FOV) when using our method versus the complex version of (Veraart et al., 2016b).

4. Discussion

We have presented a signal recovery algorithm for DWI based on singular value shrinkage of Casorati matrices with local spatial information and diffusion measures respectively arranged along the rows and the columns. Due to the limited diversity of spatial and diffusion information, particularly when correcting for unpredictable phase variations from bulk motion during the application of the diffusion gradients, it is reasonable to assume that the signal component within these matrices is approximately low rank, while the sampling noise properties can be accurately modeled when the scanner noise measures are propagated through the reconstruction. This allows the application of recent results on the asymptotic limits of additive perturbations of random matrices, most essentially the confinement of the noise bulk singular values, to obtain an objective estimate of the signal components. The method
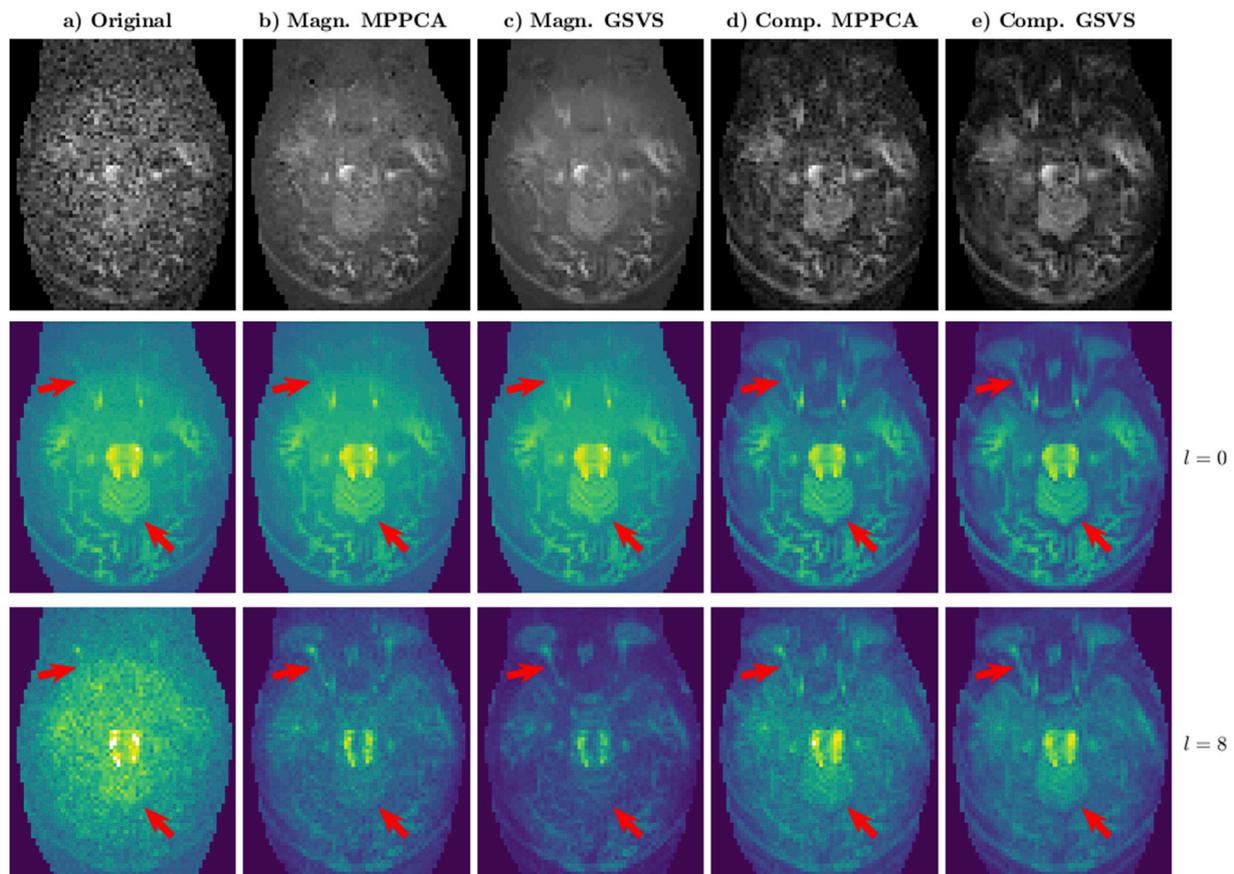
Fig. 9. a) Original dataset. **b,c)** Magnitude data denoising by **b)** MPPCA; **c)** adapting GSVS. **d,e)** Complex data denoising by **d)** adapting MPPCA; **e)** GSVS. Top: denoising results in an exemplary slice and orientation at $b = 10 \text{ms}/\mu\text{m}^2$. Mid and bottom: estimates of the spherical harmonic power for the same shell and slice respectively for orders $l = 0$ and $l = 8$. All results correspond to the optimal aspect ratio $\hat{\gamma} = 0.85$.

prevents recovery biases by using complex data information and has been applied to different brain DWI cohorts with scans obtained by advanced protocols that make use of SMS accelerated and interleaved encodings. Quantitative experiments have been conducted to validate the different assumptions and to compare to a related approach taking advantage of the estimates of the asymptotic error as provided by the theory. Our proposal guarantees optimal asymptotic risks when estimating diffusion information by singular value mappings, which stems from a synergistic combination of recent results on signal prediction using random matrix theory and the usage of the scanner noise measures.

Phase retrieval and correction has been proposed as a separate step to obtain a diffusion measure with additive noise properties, for instance, in (Prah et al., 2010; Eichner et al., 2015). These methodologies, however, inherit the limitations of the implemented phase recovery strategy, namely, potential resolution loss when low pass filtering (Prah et al., 2010), or non comprehensive use of available information when operating slicewise (Eichner et al., 2015). In contrast, our approach considers all the available information to simultaneously estimate the magnitude and the phase of the signal, so the retrieved signal accuracy is boosted by any given amount of redundancy in the complex data. Thus, although the variance of the estimates could potentially be reduced with localized phase correction, we have resorted to a simple but robust global linear correction. Our rationale is that, when compared to local phase manipulations, this correction has reduced risks of altering the statistical properties of the data and consequently biasing the complex denoising results, particularly when considering very low SNR regimes.

Patch construction in our denoising method is based on spatial proximity. However, the estimates could benefit from more refined locally adaptive criteria to build the spatial patches that also take into account the signal similarities (Katkovnik et al., 2010) as well as from patches being employed both in the spatial and the diffusion coordinates. Recent results on singular value shrinkage denoising (Leeb, 2019) show that improved estimates are possible when either the noise or the signal cannot be assumed unitarily invariant by using weighted Frobenius losses respectively after noise standardization or after an appropriate decomposition into a set of subproblems operating on subspaces of the original problem. This idea of obtaining the estimates by superposing different singular value shrinkage subproblems is actually connected with our heuristic usage of superposed patches in the image domain, so our current efforts are focused on methods to inform patch construction or problem decomposition using the theory in (Leeb, 2019).

Finally, attention should be paid to other degrading factors that may increase the rank of the matrices considered, most notably motion and non stationary distortions. In this regard, similarly to the phase correction adopted in this paper, the denoising model could include other correction steps prone to promote low-rankness while keeping the noise propagation tractability. In case that the low rank assumptions or the asymptotic regime become less justifiable, the matrix recovery problem may be reformulated using different estimation criteria (Yadav et al., 2017).

5. Conclusions

We have proposed a method for patch-based DWI retrieval based on random matrix theory. Our proposal extends previous contributions in the field by pushing the application of the theory to discontinuous, correlated and temporally heteroscedastic noise and considering more refined manipulations on the empirical singular values. Importantly, it overcomes the

limitations inherent to the simultaneous estimation of the signal and the noise properties; in our approach, the latter are modelled by the propagation of the scanner measures through the reconstruction and spectral decomposition operators. Moreover, the objective nature of this approach to denoising, using only the asymptotic properties of additive low rank perturbations of noisy matrices to find the optimal denoising strength without resorting to empirical parameter tuning, is used for patch size selection by means of the RAMSE estimates. Experiments have been conducted quantifying the benefits of noise modeling and SV shrinkage versus noise estimation and SV truncation as previously proposed, with simulations showing that our method compares favourably with related and alternative state of the art approaches. In addition, evidence has been provided on improvements when promoting low-rankness by phase demodulation and on effective operation in accelerated scanning regimes including in-plane spectral subsampling, SMS and PF as well as in temporally interleaved acquisition strategies. Finally, we have demonstrated the ability to produce denoised and debiased DWI estimates in challenging fetal, neonatal and adult neuroimaging applications.

Acknowledgments

References

Allison, M.J., Ramani, S., Fessler, J.A., 2013. Accelerated regularized estimation of MR coil sensitivities using augmented Lagrangian methods. IEEE Trans. Med. Imaging 32 (3), 556–564.

Anderson, A.W., Gore, J.C., 1994. Analysis and correction of motion artifacts in diffusion weighted imaging. Magn. Reson. Med. 32 (3), 379–387.

Baik, J., Silverstein, J.W., 2006. Eigenvalues of large sample covariance matrices of spiked population models. J. Multivar. Anal. 97 (6), 1382–1408.

Bammer, R., Holdsworth, S., Aksoy, M., Skare, S., 2010. Diffusion MRI. Oxford Univ. Press, pp. 218–249. Ch. 14: Phase errors in diffusion weighted imaging.

Basu, S., Fletcher, T., Whitaker, R., 2006. Rician noise removal in diffusion tensor MRI. In: Med. Image Comput. Comput. Assist. Interv., Part I, LNCS 4190, vol. 9, pp. 117–125. Copenhagen, Denmark.

Benaych-Georges, F., Nadakuditi, R.R., 2012. The singular values and vectors of low rank perturbations of large rectangular random matrices. J. Multivar. Anal. 111, 120–135.

Bernd, F., Schad, L.R., Klein, J., Stieltjes, B., 2009. How background noise shifts eigenvectors and increases eigenvalues in DTI. Magn. Reson. Mat. Phys. Biol. Med. 22 (3), 151–158.

Bernstein, M.A., Thomasson, D.M., Perman, W.H., 1989. Improved detectability in low signal-to-noise ratio magnetic resonance images by means of a phase-corrected real reconstruction. Med. Phys. 16 (5), 813–817.

Bhushan, C., Joshi, A.A., Leahy, R.M., Haldar, J.P., 2014. Improved B_0-distortion correction in diffusion MRI using interlaced q-space sampling and constrained reconstruction. Magn. Reson. Med. 72 (5), 1218–1232.

Bydder, M., Du, J., 2006. Noise reduction in multiple-echo data sets using singular value decomposition. Magn. Reson. Imaging 24 (7), 849–856.

Canu, S., Fourdrinier, D., 2017. Unbiased risk estimates for matrix estimation in the elliptical case. J. Multivar. Anal. 158 (4), 60–72.

Caruyer, E., 2019. Phantomas. Simulate Realistic Diffusion MRI Phantom. http://www.emmanuelcaruyer.com/phantomas.php. (Accessed 7 March 2019).

Cordero-Grande, L., 2019. MIXANDMIX: numerical techniques for the computation of empirical spectral distributions of population mixtures. Comput. Stat. & Data Anal. 1–11. https://doi.org/10.1016/j.csda.2019.06.011 in press.

Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K., 2007. Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Trans. Image Process. 16 (8), 2080–2095.

dHCP, 2019. The Developing Human Connectome Project. http://www.developingconnectome.org/. (Accessed 7 March 2019).

Dobriban, E., 2015. Efficient computation of limit spectra of sample covariance matrices. Rand. Matr. Th. Appl. 4 (4), 1–36, 1550019.

Donoho, D.L., Gavish, M., Johnstone, I.M., 2018. Optimal shrinkage of eigenvalues in the spiked covariance model. Ann. Stat. 46 (4), 1742–1778.

Edelstein, W.A., Glover, G.H., Hardy, C.J., Redington, R.W., 1986. The intrinsic signal-to-noise ratio in NMR imaging. Magn. Reson. Med. 3 (4), 604–618.

Eichner, C., Cauley, S.F., Cohen-Adad, J., Möller, H.E., Turner, R., Setsompop, K., Wald, L.L., 2015. Real diffusion-weighted MRI enabling true signal averaging and increased diffusion contrast. Neuroimage 122 (15), 373–384.

Fuderer, M., 1988. The information content of MR images. IEEE Trans. Med. Imaging 7 (4), 887–901.

Gavish, M., Donoho, D.L., 2014. The optimal hard threshold for singular values is $4/\sqrt{3}$. IEEE Trans. Inf. Theory 60 (8), 5040–5053.

Gavish, M., Donoho, D.L., 2017. Optimal shrinkage of singular values. IEEE Trans. Inf. Theory 63 (4), 2137–2152.

Gudbjartsson, H., Patz, S., 1995. The Rician distribution of noisy MRI data. Magn. Reson. Med. 34 (6), 910–914.

Hennel, F., Buehrer, M., von Deuster, C., Seuven, A., Pruessmann, K.P., 2016. SENSE reconstruction for multiband EPI including slice-dependent $N/2$ ghost correction. Magn. Reson. Med. 76 (3), 873–879.

Hutter, J., Tournier, J.D., Price, A.N., Cordero-Grande, L., Hughes, E.J., Malik, S., Steinweg, J., Bastiani, M., Sotiropoulos, S.N., Jbabdi, S., Andersson, J., Edwards, A.D., Hajnal, J.V., 2018. Time-efficient and flexible design of optimized multishell HARDI diffusion. Magn. Reson. Med. 79 (3), 1276–1292.

Jing, B.-Y., Pan, G., Shao, Q.-M., Zhou, W., 2010. Nonparametric estimate of spectral density functions of sample covariance matrices: a first step. Ann. Stat. 38 (6), 3724–3750.

Johnstone, I.M., 2001. On the distribution of the largest eigenvalue in principal component analysis. Ann. Stat. 29 (2), 295–327.

Jones, D.K., Basser, P.J., 2004. "Squashing peanuts and smashing pumpkins": how noise distorts diffusion-weighted MR data. Magn. Reson. Med. 52 (5), 979–993.

Kammoun, A., Alouini, M.-S., 2016. No eigenvalues outside the limiting support of generally correlated Gaussian matrices. IEEE Trans. Inf. Theory 62 (7), 4312–4326.

Katkovnik, V., Foi, A., Egiazarian, K., Astola, J., 2010. From local kernel to nonlocal multiple-model image denoising. Int. J. Comput. Vis. 86 (1), 1–32.

Kellman, P., McVeigh, E.R., 2005. Image reconstruction in SNR units: a general method for SNR measurement. Magn. Reson. Med. 54 (6), 1439–1447.

Koay, C.G., Basser, P.J., 2006. Analytically exact correction scheme for signal extraction from noisy magnitude MR signals. J. Magn. Reson. 179 (2), 317–322.

Ledoit, O., Wolf, M., 2017. Numerical implementation of the QuEST function. Comput. Stat. Data Anal. 115, 199–223.

Leeb, W., 2019. Matrix denoising for weighted loss functions and heterogeneous signals, pp. 1–29 arXiv:1902.09474v1. https://arxiv.org/abs/1902.09474.

Manjón, J.V., Coupé, P., Concha, L., Buades, A., Collins, D.L., Robles, M., 2013. Diffusion weighted image denoising using overcomplete local PCA. PLoS One 8 (9), 205–218.

Marčenko, V.A., Pastur, L.A., 1967. Distribution of eigenvalues for some sets of random matrices. Math. USSR Sbornik 1 (4), 457–483.

Nadakuditi, R.R., 2014. Optshrink: an algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage. IEEE Trans. Inf. Theory 60 (5), 3002–3018.

Noll, D.C., Nishimura, D.G., Macovski, A., 1991. Homodyne detection in magnetic resonance imaging. IEEE Trans. Med. Imaging 10 (2), 154–163.

Pai, V.M., Rapacchi, S., Kellman, P., Croisille, P., Wen, H., 2011. PCATMIP: enhancing signal intensity in diffusion-weighted magnetic resonance imaging. Magn. Reson. Med. 65 (6), 1611–1619.

Prah, D.E., Paulson, E.S., Nencka, A.S., Schmainda, K.M., 2010. A simple method for rectified noise floor suppression: phase-corrected real data reconstruction with application to diffusion-weighted imaging. Magn. Reson. Med. 64 (2), 418–429.

Pruessmann, K.P., Weiger, M., Scheidegger, M.B., Boesiger, P., 1999. SENSE: sensitivity encoding for fast MRI. Magn. Reson. Med. 42 (5), 952–962.

Pruessmann, K.P., Weiger, M., Börtner, P., Boesiger, P., 2001. Advances in sensitivity encoding with arbitrary k-space trajectories. Magn. Reson. Med. 46 (4), 638–651.

Setsompop, K., Cohen-Adad, J., Gagoski, B.A., Raij, T., Yendiki, A., Keil, B., Wedeen, V.J., Wald, L.L., 2012. Improving diffusion MRI using simultaneous multi-slice echo planar imaging. Neuroimage 63 (1), 569–580.

Silverstein, J.W., 1995. Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. J. Multivar. Anal. 55 (2), 331–339.

St-Jean, S., Coupé, P., Descoteaux, M., 2016. Non local spatial and angular matching: enabling higher spatial resolution diffusion MRI datasets through adaptive denoising. Med. Image Anal. 32, 115–130.

Tournier, J.-D., Smith, R., Raffelt, D., Tabbara, R., Dhollander, T., Pietsch, M., Christiaens, D., Jeurissen, B., Yeh, C.-H., Connelly, A., 2019. Connelly, Mrtrix3: a fast, flexible and open software framework for medical image processing and visualization. bioRxiv 551739, 1–27. https://doi.org/10.1101/551739.

Tristán-Vega, A., Aja-Fernández, S., 2010. DWI filtering using joint information for DTI and HARDI. Med. Image Anal. 14 (2), 205–218.

Veraart, J., Fieremans, E., Novikov, D.S., 2016. Diffusion MRI noise mapping using random matrix theory. Magn. Reson. Med. 76 (5), 1582–1593.

Veraart, J., Novikov, D.S., Christiaens, D., Ades-aron, B., Sijbers, J., Fieremans, E., 2016. Denoising of diffusion MRI using random matrix theory. Neuroimage 142 (15), 394–406.

Wagner, S., Couillet, R., Debbaha, M., Slock, D.T.M., 2012. Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback. IEEE Trans. Inf. Theory 58 (7), 4509–4537.

Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 13 (4), 600–612.

Wirestam, R., Bibic, A., Lätt, J., Brockstedt, S., Ståhlberg, F., 2006. Denoising of complex MRI data by wavelet-domain filtering: application to high-b -value diffusion-weighted imaging. Magn. Reson. Med. 56 (5), 1114–1120.

Wood, J.C., Johnson, K.M., 1999. Wavelet packet denoising of magnetic resonance images: importance of Rician noise at low SNR. Magn. Reson. Med. 41 (3), 631–635.

Yadav, S.K., Sinha, R., Bora, P.K., 2017. A fast data driven shrinkage of singular values for arbitrary rank signal matrix denoising, pp. 1–13 arXiv:1701.05223v2. https://arxiv.org/abs/1701.05223.

Zhu, K., Dougherty, R.F., Wu, H., Middione, M.J., Takahashi, A.M., Zhang, T., Pauly, J.M., Kerr, A.B., 2016. Hybrid-space SENSE reconstruction for simultaneous multi-slice MRI. IEEE Trans. Med. Imaging 35 (8), 1824–1836.