Received: 4 April 2022 | Revised: 17 August 2022 | Accepted: 15 September 2022 | Published on: 11 October 2022

DOI: 10.1002/mrm.29478

RESEARCH ARTICLE

Magnetic Resonance in Medicine

Tensor denoising of multidimensional MRI data

Jonas L. Olesen[1,2] | **Andrada Ianus**[3] | **Leif Østergaard**[1] | **Noam Shemesh**[3] | **Sune N. Jespersen**[1,2]

[1]Center of Functionally Integrative Neuroscience, Department of Clinical Medicine, Aarhus University, Aarhus, Denmark

[2]Department of Physics and Astronomy, Aarhus University, Aarhus, Denmark

[3]Champalimaud Research, Champalimaud Foundation, Lisbon, Portugal

Correspondence
Center of Functionally Integrative Neuroscience, Department of Clinical Medicine, Aarhus University, 8000 Aarhus C, Denmark.
Email: sune@cfin.au.dk

Funding information
"la Caixa" Foundation, Grant/Award Number: 100010434; Danish Ministry of Science, Innovation, and Education, Grant/Award Number: MINDLab; Danish National Research Foundation, Grant/Award Number: CFIN; European Regional Development Fund, Grant/Award Number: LISBOA-01-0145-FEDER-02217; Fundação para a Ciência e Tecnologia; H2020 European Research Council, Grant/Award Numbers: 100010434, agreement 679058; Lisboa Regional Operational Programme, Grant/Award Number: LISBOA-01-0145-FEDER-02217; Velux Fonden, Grant/Award Numbers: ARCADIA, grant 00015963

Purpose: To develop a denoising strategy leveraging redundancy in high-dimensional data.

Theory and Methods: The SNR fundamentally limits the information accessible by MRI. This limitation has been addressed by a host of denoising techniques, recently including the so-called MPPCA: principal component analysis of the signal followed by automated rank estimation, exploiting the Marchenko-Pastur distribution of noise singular values. Operating on matrices comprised of data patches, this popular approach objectively identifies noise components and, ideally, allows noise to be removed without introducing artifacts such as image blurring, or nonlocal averaging. The MPPCA rank estimation, however, relies on a large number of noise singular values relative to the number of signal components to avoid such ill effects. This condition is unlikely to be met when data patches and therefore matrices are small, for example due to spatially varying noise. Here, we introduce tensor MPPCA (tMPPCA) for the purpose of denoising multidimensional data, such as from multicontrast acquisitions. Rather than combining dimensions in matrices, tMPPCA uses each dimension of the multidimensional data's inherent tensor-structure to better characterize noise, and to recursively estimate signal components.

Results: Relative to matrix-based MPPCA, tMPPCA requires no additional assumptions, and comparing the two in a numerical phantom and a multi-TE diffusion MRI data set, tMPPCA dramatically improves denoising performance. This is particularly true for small data patches, suggesting that tMPPCA can be especially beneficial in such cases.

Conclusions: The MPPCA denoising technique can be extended to high-dimensional data with improved performance for smaller patch sizes.

KEYWORDS
denoising, diffusion, principal component analysis, random matrix theory

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2022 The Authors. *Magnetic Resonance in Medicine* published by Wiley Periodicals LLC on behalf of International Society for Magnetic Resonance in Medicine.

15222594, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mrm.29478 by Cochrane Portugal, Wiley Online Library on [16/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

1 | INTRODUCTION

Due to its noninvasive nature, MRI is used widely in diagnostics and biomedical research. The SNR of MRI is inherently low, however, leading to lower spatiotemporal-resolution images or long acquisition times, both of which are undesirable for the subjects, and particularly vulnerable patients, undergoing an MRI session. While this applies to all MRI methods, some are more affected by SNR limitations, as they require active signal attenuation to produce contrast. For example, while diffusion MRI (dMRI)[1] holds great promise as a technique for interrogating tissue microstructure,[2] it is especially limited by noise, because the signal is strongly attenuated by the application of diffusion gradients.[3] Noise thus constitutes a major limitation for inference whether by qualitative visual inspection or quantitative modeling, and a source that cannot be expected to be resolved by hardware innovation alone as current hardware already achieves approximately 90% of the ultimate intrinsic SNR.[4–6] Note that increasing the SNR either by increases in magnetic field or by using cryogenic coils[7] comes at a relatively high cost and may still benefit from denoising to further extend their reach.

Denoising methods can partly alleviate this SNR limitation and therefore can constitute an important step in MRI data analysis. Various denoising methods have been adopted within the MRI community, notably nonlocal means filtering,[8–10] total variation minimization,[11] discrete cosine transform filtering,[10] neural networks,[12,13] and local principal component analysis (PCA).[14,15] Local PCA divides the MRI data set into local patches, which are projected into a low-rank approximation given by the most significant principal components. The MRI signal is typically well-described by a few such components,[16] while thermal noise will be evenly distributed among all the components, thereby enabling noise reduction in proportion to the square root of the number of discarded components. The number of significant signal components can be estimated objectively by exploiting that noise singular values are distributed according to the Marchenko-Pastur (MP) distribution.[17] While being a seasoned idea in MRI as such,[18,19] it was only recently introduced to dMRI and coupled with local PCA as MPPCA.[15,16] The latter can be seen as a more "objective" approach to denoising, making its application easier relative to earlier approaches that had to resort to choosing or empirically estimating the number of signal components in a more ad hoc manner. In particular, MPPCA was shown to compare favorably with other state-of-the-art denoising techniques in terms of noise reduction without compromising anatomical detail or blurring the data.[15]

Because of its objective rank estimation and effectiveness, MPPCA has become a very popular SNR-boosting technique.[20–26] The implementation introduced by Veraart et al[15,16] shapes the data within each patch into a voxel × modality matrix. The current formulation of the objective MP rank estimation relies on the number of signal components being small relative to the matrix dimensions, as the used distribution of noise singular values only formally applies in the limit of infinitely large matrices and absence of signal components. This is problematic in practical applications in which the denoised matrix is not very large, such as scenarios with small patches necessitated by, for example, spatially varying noise.

Here, we present a revised distribution that is applicable even when the number of signal components cannot be considered small. We thereby extend the applicability of MPPCA. Building on the success of MPPCA, we further propose a generalization to multidimensional data, such as arising with multicontrast acquisitions. The structure of such data sets is not used by standard MPPCA, which partially discards the information in the data's structure by reshaping it into a matrix. We therefore propose tensor MPPCA (tMPPCA), which retains the patches' natural tensor structure and improves the noise reduction by using the additional redundancy available in the extra dimensions. To achieve this, tMPPCA builds on higher-order singular value decomposition (HOSVD), which has previously been used in various forms for denoising dMRI and MRI more generally.[27–32] Because HOSVD consists of a sequence of regular SVDs, we can exploit that objective MP rank estimation is applicable to each dimension separately, thereby extending the application of MPPCA to multidimensional data. The denoising performance of tMPPCA is further enhanced relative to regular HOSVD approaches by recursive estimation of the signal components (ie, reducing the noise for one dimension improves the signal estimation for the other dimensions).

We compare the performances of tMPPCA and MPPCA in multi-echo diffusion data as an example of multidimensional data, and in simulations. We find that tensor MPPCA performs especially well relative to MPPCA for small patch sizes.

2 | THEORY

The objective rank estimation in MPPCA is facilitated by the MP distribution of noise singular values as detailed by Veraart et al.[15] Here, we reiterate the approach and modify the distribution of noise singular values to take the effect of a nonzero number of signal components into account.

15222594, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mrm.29478 by Cochrane Portugal, Wiley Online Library on [16/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

In essence, the approach considers an $M \times N$ matrix X with singular value decomposition (SVD)

$$X = USV^{\dagger} \tag{1}$$

where U and V are unitary matrices, and S is a diagonal matrix containing the singular values. Let the P largest singular values represent signal components, while the smaller singular values are due to zero mean independent and identically distributed noise.

Consider first the case $P = 0$, and define $N' \equiv \max(M, N)$ and $M' \equiv \min(M, N)$. Assuming X to be full rank due to noise, M' is then the number of nonzero singular values. In the following, "singular values" will refer to the nonzero singular values. Each singular value S_{ii} is associated with an eigenvalue $\lambda_i = S_{ii}^2$ of the "covariance matrix" $XX^{\dagger}$. The eigenvalues approximately follow the MP distribution. Here, we write the MP distribution rescaled to account for the omitted $1/\sqrt{N}$ and modified to exclude zero-valued singular values; we will still refer to it as the MP distribution:

$$p(\lambda) = \begin{cases} \frac{\sqrt{(\lambda - \lambda_-)(\lambda_+ - \lambda)}}{2\pi\sigma^2 M' \lambda}, & \lambda_- < \lambda < \lambda_+ \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

$$\lambda_{\pm} = \sigma^2 \left(\sqrt{N'} \pm \sqrt{M'} \right)^2 \tag{3}$$

Formally, the distribution only applies in the limit $M, N \to \infty$ with constant ratio M/N.[17] A finite size of X introduces tails in the sense that eigenvalues can exceed the otherwise hard upper and lower bounds $\lambda_{\pm}$ as exemplified in Figure 1. The distribution is also affected when $P > 0$. However, the MP distribution can be heuristically generalized to nonzero P simply by subtracting P from M' and N' (Figure 1). This can be verified explicitly with Monte Carlo simulations. Furthermore, Figure 2, which shows simulations for the full range of M/N and up to 80% signal ratio P/M, compares the eigenvalues' mean and upper bound as predicted by the modified distribution to demonstrate the applicability of the modified distribution in virtually any practical scenario.

The mean of the MP distribution is $\lambda = N'\sigma^2$. Accordingly, when taking $P \geq 0$ into account, the average sum of the noise eigenvalues is

$$\sum_i \lambda_i = \left(M' - P\right)\left(N' - P\right)\sigma^2. \tag{4}$$

This property suggests the following algorithm for determining σ^2 and P (for λ_i sorted in descending order):

For $P = 0, 1, \ldots$

- Calculate

$$\sigma_P^2 = \frac{1}{(M' - P)(N' - P)} \sum_{i=P+1}^{M'} \lambda_i$$

- Calculate

$$\lambda_+^{(P)} = \sigma_P^2 \left(\sqrt{N'} + \sqrt{M'} \right)^2$$

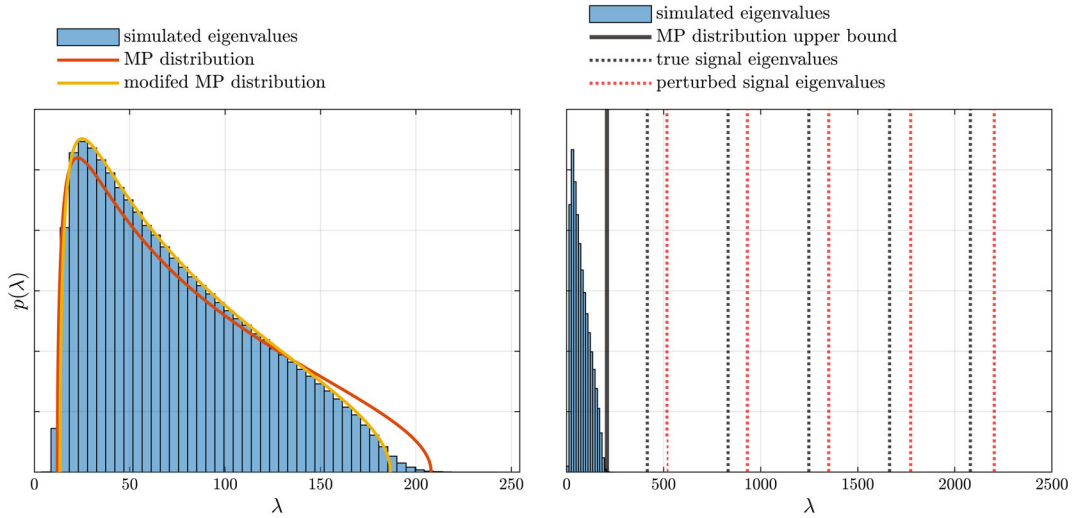

FIGURE 1 Distribution of squared singular values of a $M = 30$, $N = 80$ matrix with $\sigma^2 = 1$ Gaussian noise and $P = 5$ signal components with values given in the plots as black dotted vertical lines. Without loss of generality, the matrix was generated on the basis of the signal eigenvectors. The distribution of noise eigenvalues and mean perturbed signal eigenvalues were calculated from 10^5 noise realizations. The left panel shows the simulated distribution compared with the Marchenko-Pastur (MP) distribution given in Equation (2) and the MP distribution modified by subtracting P from M' and N'. In the right panel, the x-range is increased to encompass the true signal eigenvalues as well as the perturbed mean values of the corresponding simulated eigenvalues

15222594, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mrm.29478 by Cochrane Portugal, Wiley Online Library on [16/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

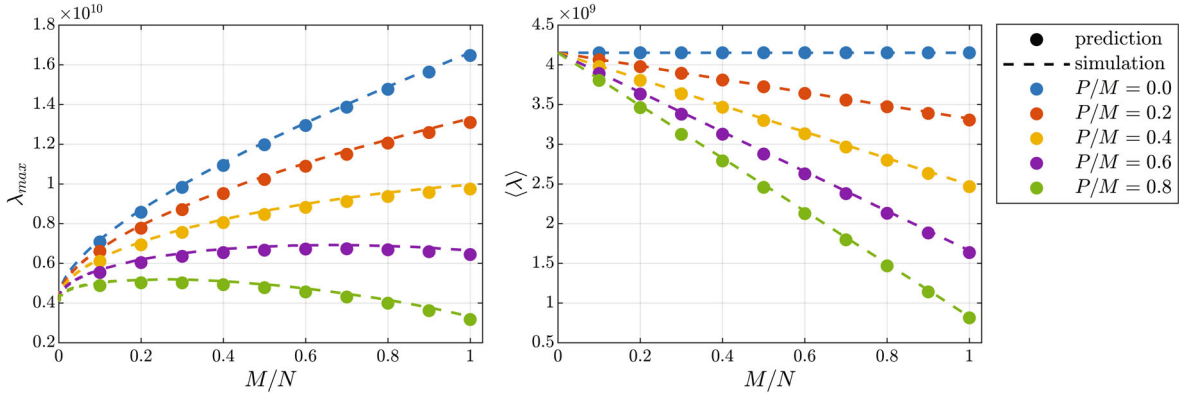

FIGURE 2 Prediction and simulation of the largest and mean noise eigenvalue (squared singular value) when X contains P signal components. The prediction is according to the modified MP distribution (subtracting P from M' and N' in Equation [2]), and the simulated values were generated using 10^3 Gaussian noise ($\sigma^2 = 1$) realizations added to a matrix with signal eigenvalues equaling 10 times the upper bound λ_+ of the MP distribution. The value of $N = 10^3$ is fixed, while the ratio M/N is varied approximately over its entire relevant range from 0 to 1. Each curve is associated with a fixed value for the ratio of signal components P/M

- If $\lambda_{P+1} < \lambda_+^{(P)}$, terminate and output the current value for P and $\sigma^2 \approx \sigma_P^2$

In words, remove apparent signal singular values from the estimate of σ^2 until there is self-consistency (ie, all singular values lie below the predicted MP upper bound). While the modified MP distribution yields an accurate estimate of σ^2, we found that it works best in practice to use the upper bound λ_+ from the unmodified MP distribution. This overestimates λ_+ and consequently includes less of the noise distribution's upper tail as apparent signal components (see Figure 1).

After removing the $M' - P$ noise components, some further improvement is attainable by applying optimal shrinkage,[21,33] which approximately corrects the noise perturbation of the signal eigenvalues, illustrated in Figure 1.

We emphasize that the MP approach cannot detect signal components below the upper bound of the noise distribution λ_+ and will consequently remove them. However, this is typically unproblematic in practice, because most of the signal variance usually resides in a few large components. On the other hand, it poses a problem in scenarios with very low SNR. Also of note, from the perspective of the MP-based denoising, data artifacts such as k-space spikes are signal on equal footing with the "true" signal and will be retained. Such artifacts can therefore limit the denoising performance of MPPCA-based approaches by increasing the number of apparent signal components.

2.1 | Marchenko-Pastur PCA denoising

Conventional MPPCA denoising[15,16] applies a sliding window over voxels and denoises the resulting patches independently. Each patch forms a matrix X with voxels as rows and one or more modalities (eg, diffusion, relaxation, functional MRI) as columns. Following SVD of X, the noise components are removed based on this approach. This can be done by projecting X onto the signal subspace as follows:

$$\widetilde{X} = \widetilde{U}\widetilde{U}^\dagger X = \widetilde{U}\widetilde{S}\widetilde{V}^\dagger \tag{5}$$

where $\widetilde{U}$ is U with the $M - P$ columns associated with noise eigenvalues removed; U is an $M \times M'$ matrix; and $\widetilde{U}$ is an $M \times P$ matrix. Similarly, $\widetilde{S}$ is a $P \times P$ diagonal matrix and $\widetilde{V}$ is an $N \times P$ matrix.

The total noise variance in X is $MN\sigma^2$, while the removed variance is given by Equation (4). The noise variance after denoising is their difference, yielding a resulting noise variance of

$$\widetilde{\sigma}^2 \approx \frac{P^2 + (M-P)P + (N-P)P}{MN}\sigma^2 \tag{6}$$

The denoised patches can be combined in one of several ways.[15] The best performance is achieved by having a sliding stride of one voxel and averaging the patches[14,34] (ie, each voxel receives an equal contribution from each patch of which it is a member). Alternatively, one can assign the denoised signal only to the center voxel in each patch. Notably, the remaining noise will be correlated among voxels in both cases, which is potentially problematic in some applications. If desired, this can be avoided at the cost of decreased denoising efficiency by combining all voxels in one patch.

2.2 | Higher-order SVD denoising

Our proposed tMPPCA denoising is similar to denoising with HOSVD with some key differences. Here

15222594, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mrm.29478 by Cochrane Portugal, Wiley Online Library on [16/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

we outline the HOSVD approach to highlight these differences.

For multidimensional data, each patch has the natural structure of a tensor, and X can be decomposed using HOSVD. This can be done by applying regular SVD to each flattening of X, where flattening refers to the matrix formed by concatenating all but one index. Assume without loss of generality that each patch is a three-index tensor $X_{i_1 i_2 i_3}$ (for instance, in dMRI with i_1 for voxels, i_2 for directions, and i_3 for b-values with X being an $M_1 \times M_2 \times M_3$ tensor). The i_1-flattening of tensor X is the $M_1 \times M_2 M_3$ matrix $X_{i_1(i_2 i_3)}$, where $(\dots)$ signifies index concatenation. In this way, there is one flattening and associated SVD for each index. Assume the number of signal components is known for each flattening. Then, a denoised tensor can be constructed analogously to Equation (5).

$$\widetilde{X} = \left(\widetilde{U}_1\widetilde{U}_1^\dagger, \widetilde{U}_2\widetilde{U}_2^\dagger, \widetilde{U}_3\widetilde{U}_3^\dagger\right) \cdot X = \left(\widetilde{U}_1, \widetilde{U}_2, \widetilde{U}_3\right) \cdot \widetilde{S} \quad (7)$$

Where $\cdot$ denotes tensor multiplication in the sense that $\widetilde{U}_n$ acts on the nth index. The tensor $\widetilde{S} = \left(\widetilde{U}_1^\dagger, \widetilde{U}_2^\dagger, \widetilde{U}_3^\dagger\right) \cdot X$ with dimensions $P_1 \times P_2 \times P_3$ is the "core," which generalizes the diagonal matrix from SVD. For clarity, note that if X is a matrix (rank-2 tensor), then the first flattening is simply X while the second flattening is X transposed, so $U_1 = U$ and $U_2 = V^\dagger$ in this notation.

As with regular SVD denoising, denoising tensors with HOSVD has been performed using an a priori choice for the signal rank (for each dimension).[30] However, we point out that because HOSVD amounts to a sequence of regular SVDs, objective MP-based rank estimation is applicable to each flattening. Thereby, MPPCA can be applied to tensors while exploiting the redundancy of the additional dimensions. For further improvement, we introduce a recursive step, which then defines the proposed tMPPCA denoising as detailed subsequently.

2.3 | Tensor MPPCA denoising

The proposed tMPPCA denoising modifies the HOSVD approach as follows. Consider X's i_1-flattening $X_{i_1(i_2 i_3)}$ and the associated SVD, $U_1 S_1 V_1^\dagger$. Now, MP rank reduction is applicable and denoises the flattening to $\widetilde{U}_1\widetilde{S}_1\widetilde{V}_1^\dagger$. Define the tensor $\widetilde{X}_1 \equiv \widetilde{U}_1^\dagger \cdot_{i_1} X$, where $\cdot_{i_n}$ signifies contraction of index i_n. While X has dimensions $M_1 \times M_2 \times M_3$, $\widetilde{X}_1$ has dimensions $P_1 \times M_2 \times M_3$, where P_1 is the number of identified signal components in the MP rank-reduction step. Crucially, $\widetilde{X}_1$ retains X's noise properties because $\widetilde{U}_1$ is (semi-) unitary. Therefore, the procedure can be repeated by applying SVD and rank reduction to the i_2-flattening of $\widetilde{X}_1$, yielding $\widetilde{U}_2$ and $\widetilde{X}_2 \equiv \widetilde{U}_2^\dagger \cdot_{i_2} \widetilde{X}_1$ and so on. The recursive procedure stops at the last index, after which a denoised version of X can be reconstructed using Equation (7) in the same way as for HOSVD.

Note that the SVD at step n depends on the denoising at step $n-1$, which improves the estimates of $\widetilde{U}_n$ and $\widetilde{X}_n$. In contrast, each U_n is estimated independently in the HOSVD approach. Therefore, the ordering of the indices/flattenings does not matter for HOSVD but does matter for tMPPCA. Choosing the index ordering therefore constitutes an additional input for tMPPCA, which potentially influences the performance appreciably if the ratio of signal components varies sufficiently between flattenings. For instance, for the demonstration data set used here (six b-shells), we found that using the b-value index for the initial flattening was suboptimal due to the resultantly small number of singular values. One method for choosing the ordering is given in section 3.1. In contrast to HOSVD, optimal shrinkage is applicable to the final set of signal eigenvalues $\widetilde{S}_n$, further improving performance.

Each SVD followed by MP rank reduction produces an estimate of σ^2. Because the noise properties are unchanged throughout the procedure, these estimates are ideally equal but will differ slightly due to finite dimensions and random noise. Therefore, the σ^2 estimate can be improved by combining the individual estimates in a weighted sum. Similar to inverse-variance weighting, we weigh the σ^2 estimates by their distributions' widths $(\lambda_+ - \lambda_-)^2 \propto (M-P)(N-P)$. As such, tMPPCA improves the noise characterization in addition to the denoising itself. However, it requires an initial pass in which the SVD of each flattening is calculated before any denoising.

Tensor MPPCA reduces to MPPCA when X is a matrix, thereby providing equal performance when no additional redundancy is available but improved performance otherwise. It outperforms classical HOSVD approaches due to its objective rank estimation and the recursive approach, which improves each subsequent SVD and enables the application of optimal shrinkage. A conceptual overview (not an efficient implementation) of the algorithm is given here, assuming the indices of X have been ordered as desired.

- (Calculate improved σ^2 estimate from noise singular values of all flattenings)
- Set $\widetilde{X}_0 = X$
- For $n = 1, 2, \dots, \eta$ (η = number of indices in X)
  - Calculate the SVD $U_n S_n V_n^\dagger$ of the i_n flattening of $\widetilde{X}_{n-1}$
  - Apply MP rank reduction, yielding $\widetilde{U}_n\widetilde{S}_n\widetilde{V}_n^\dagger$
  - Set $\widetilde{X}_n = \widetilde{U}_n^\dagger \cdot_{i_n} \widetilde{X}_{n-1}$

15222594, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mrm.29478 by Cochrane Portugal, Wiley Online Library on [16/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

- (Apply optimal shrinkage to $\widetilde{S}_\eta$, modifying $\widetilde{X}_\eta$ accordingly; the i_η-flattening of $\widetilde{X}_\eta$ is $\widetilde{S}_\eta \widetilde{V}_\eta^\dagger$)
- Output $\widetilde{X} = \left(\widetilde{U}_1, \cdots, \widetilde{U}_\eta\right) \cdot \widetilde{X}_\eta$

Subtracting the removed variance at each step of tMPPCA from the total noise variance $M_1 \cdots M_\eta \sigma^2$ (same approach as for MPPCA) yields the average noise variance after denoising

$$\widetilde{\sigma}^2 \approx \frac{P_1 \cdots P_\eta + \sum_n (M_n - P_n) P_n}{M_1 \cdots M_\eta} \sigma^2 \tag{8}$$

3 | METHODS

3.1 | Denoising

We provide a *MATLAB* implementation of tMPPCA openly available at https://github.com/sunenj/MP-PCA-Denoising.

We compare the performance of matrix and tMPPCA implemented according to this theory (ie, including the revised MP distribution). This improves the performance in specific scenarios but will typically only result in subtle differences, and this is also the case for the example data sets used here. Using the revised MP distribution matters more for tMPPCA, as this approach is more likely to encounter noisy matrices with a relatively large number of signal components because the resulting matrix after denoising in each index contains progressively less noise. In practice, we can then use the tMPPCA implementation for both tensor and MPPCA. We emphasize that also in the case of regular MPPCA, the entire data set is used in the denoising algorithm by concatenating the echo and diffusion dimensions. In the specific demonstration data presented here, the alternative would be applying it independently for each TE.

For tMPPCA specifically, we always choose the index ordering by maximizing the amount of retained signal variance. This is suboptimal in terms of noise reduction, which is generally maximized by discarding the maximal number of components. However, it is the most conservative approach in terms of minimizing the amount of discarded true signal variance. The idea is that the amount of true signal variance above the upper bound of the MP distribution might vary between the flattenings of X and can be estimated as the sum of squared signal singular values for each flattening. The signal singular values are inflated by noise, and this was corrected for using optimal shrinkage.[33] In principle, the optimal index ordering according to this criterion can vary between patches. However, at least for the data sets used here, we found that the optimal ordering either did not vary between patches, or if it did vary, the retained signal variances were practically equal between the relevant indices. Therefore, for simplicity, the same ordering was used for all patches. Each ordering used here is specified in the caption of the corresponding figure.

3.2 | Data

All experiments were preapproved by the competent institutional and national authorities and were carried out in accordance with European Directive 2010/63.

To demonstrate the benefit of tMPPCA on data sets with multiple redundancy sources, we use multi-TE diffusion data with a four-index structure (voxels $\times\, \hat{g} \times b \times$ TE). Several images were below the noise floor due to combined strong diffusion weighting and long TE, but we recover MR images with sufficient SNR for applying diffusion kurtosis imaging (DKI).

The acquisition was performed on a mouse brain ex vivo. Following transcardial perfusion, the brain was removed, immersed in 4% paraformaldehyde solution (24 h), and washed in phosphate-buffered saline solution (48 h). Then, the sample was placed in a 10-mm NMR tube with Fluorinert kept at 22°C and scanned using a 16.4T Bruker Eon scanner with a Micro5 probe.

The data were acquired using a diffusion-weighted RARE sequence (remmiRARE from https://remmi-toolbox.github.io/) with FOV = $10 \times 8\,\text{mm}^2$, matrix size = 100×80, in-plane resolution $100 \times 100\,\mu\text{m}^2$, slice thickness = 500 μm, and 27 slices. The data were acquired with partial Fourier = 1.2 in the phase direction with resultant matrix size = 100×67. Other parameters were TR = 4 s, gradient pulse width δ = 3.5 ms, gradient pulse separation Δ = 5.6 ms, 20 gradient directions uniformly distributed on a hemisphere, and six b-values linearly distributed from 0.5 to 3 ms/μm^2. The number of TEs was 38 with TE linearly varied between 11 and 111 ms. The total scan time was 10 h.

The raw k-space data were Fourier-transformed and denoised with either MPPCA or tMPPCA using a 10×10 sliding window and patch averaging (other window sizes were used for the Supporting Information and are specified there). Before denoising, a spike artifact specific to the used data set was partially corrected by replacing the relevant patch of k-space data with a similar patch from the corner of acquired k-space. This substantially suppressed the artifact, but subtle vertical lines are visible in some images. Due to the use of partial Fourier, the effective in-plane resolution of the data is reduced to $100 \times 84\,\mu\text{m}^2$. Because the denoising was performed on the raw data, it was unnecessary to recover the full resolution for demonstrating the denoising performance, and this processing

15222594, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mrm.29478 by Cochrane Portugal, Wiley Online Library on [16/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

step was therefore skipped. To further reduce memory and compute intensity, only a single slice was used. However, the presented results and conclusions generalize to multi-slice data with the ideal window depth (potentially single slice) being dependent on the slice thickness and specific tissue just as the optimal 2D window size depends on similar factors. Magnitude data were used in the subsequent DKI analysis.

3.3 | Simulations

The denoising performance was also assessed using a numerical phantom. This provides the benefit of a known ground truth. The phantom was created by using the fitted DKI parameters from the denoised multi-TE diffusion data set to generate a DKI signal for each TE. The TE values for the phantom are therefore identical to those acquired experimentally. However, only the 20 smallest TE values (up to TE = 62 ms) were used to ensure very high SNR for the DKI fit. The same *b*-values and number of gradient directions were used as in the original data set except when stated otherwise. Gaussian noise of SNR $\equiv \frac{S_0}{\sigma} = 20$ was added to the phantom data before denoising, and the denoising performance is quantified as the RMS error between the denoised and ground-truth data. This performance metric reflects the denoising algorithm's ability to recover the actual ground truth rather than simply the amount of removed variance, which potentially includes lost true signal variance.

4 | RESULTS

4.1 | Multi-TE diffusion data

The denoising performances of MPPCA and tMPPCA on the acquired multi-TE data are compared in Figure 3. Tensor MPPCA benefits considerably from the redundancy of the multiple TEs, achieving an estimated 3-fold SNR gain relative to MPPCA. Because the ground truth is unknown, the SNR gains are estimated using Equations (4) and (8), respectively. At small TE, the MPPCA and tMPPCA denoised images are indistinguishable, but at large TE, the improved quality of the images recovered using tMPPCA is apparent. Notably, in the regular MPPCA denoising, many features of the brain at long TE are lost, but tMPPCA clearly better preserves the information in terms of agreement with expected tissue symmetries and features. The noise reduction is appreciated by comparing single-voxel signals (Figure 3C), where the expected monotonous signal decay can be appreciated only from the tensor denoising. Figure 3 also includes residual distributions, demonstrating lack of structure and excellent agreement with Gaussian statistics, and suggesting that both MPPCA and tMPPCA remove a negligible amount of true signal variance relative to the amount of removed Gaussian noise.

This example data set consists of a large number of images with negligible spatial variation in the noise variance. Accordingly, the performance of MPPCA is improved by using a larger sliding window. In the limiting case of denoising all voxels as one patch, MPPCA achieves a factor of 2.5 SNR gain relative to the 10×10 window used when producing Figure 3 (see Supporting Information Figure S1), whereas the performance of tMPPCA decreases and becomes comparable but still larger. The MPPCA technique then reproduces the tissue symmetries and features at large TE otherwise only visible for tMPPCA in Figure 3A. This provides verification of the denoised images from tMPPCA (at least with MPPCA regarded as a standard).

In contrast, Supporting Information Figure S2 shows the contents of Figure 3 produced using a 3×3 sliding window. The performances of matrix and tMPPCA are both decreased, and slight deviations from Gaussian residuals are observed for MPPCA, as can be expected due to violating the large matrix assumption in the MP rank reduction. Tensor MPPCA is less affected, as it can rely on redundancy in the other indices and maintains good performance despite the small window size.

Figure 4 compares DKI parameter maps for the subset of denoised data with smallest and largest TE. Again, the maps show little differences at the smallest TE where the SNR is high, but at the largest TE they are distorted substantially when using data denoised with MPPCA. This is mediated to a large degree by using the data denoised with tMPPCA.

4.2 | Simulations

Figure 5 compares the performance of MPPCA and tMPPCA applied to simulated multi-TE diffusion data with varying number of TEs, gradient directions, and varying window size and patch method. Comparing the performances at varying number of TEs and gradient directions shows that tMPPCA benefits substantially from the additional redundancy this introduces. In comparison, the performance of MPPCA plateaus early. Tensor MPPCA outperforms MPPCA for all window sizes, consistent with MPPCA being a special case.

Generally, the performance increases with window size until an optimum, which depends on the data-set size, patch combination method, and SNR. Tensor MPPCA reaches optimal performance at considerably smaller

15222594, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mrm.29478 by Cochrane Portugal, Wiley Online Library on [16/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

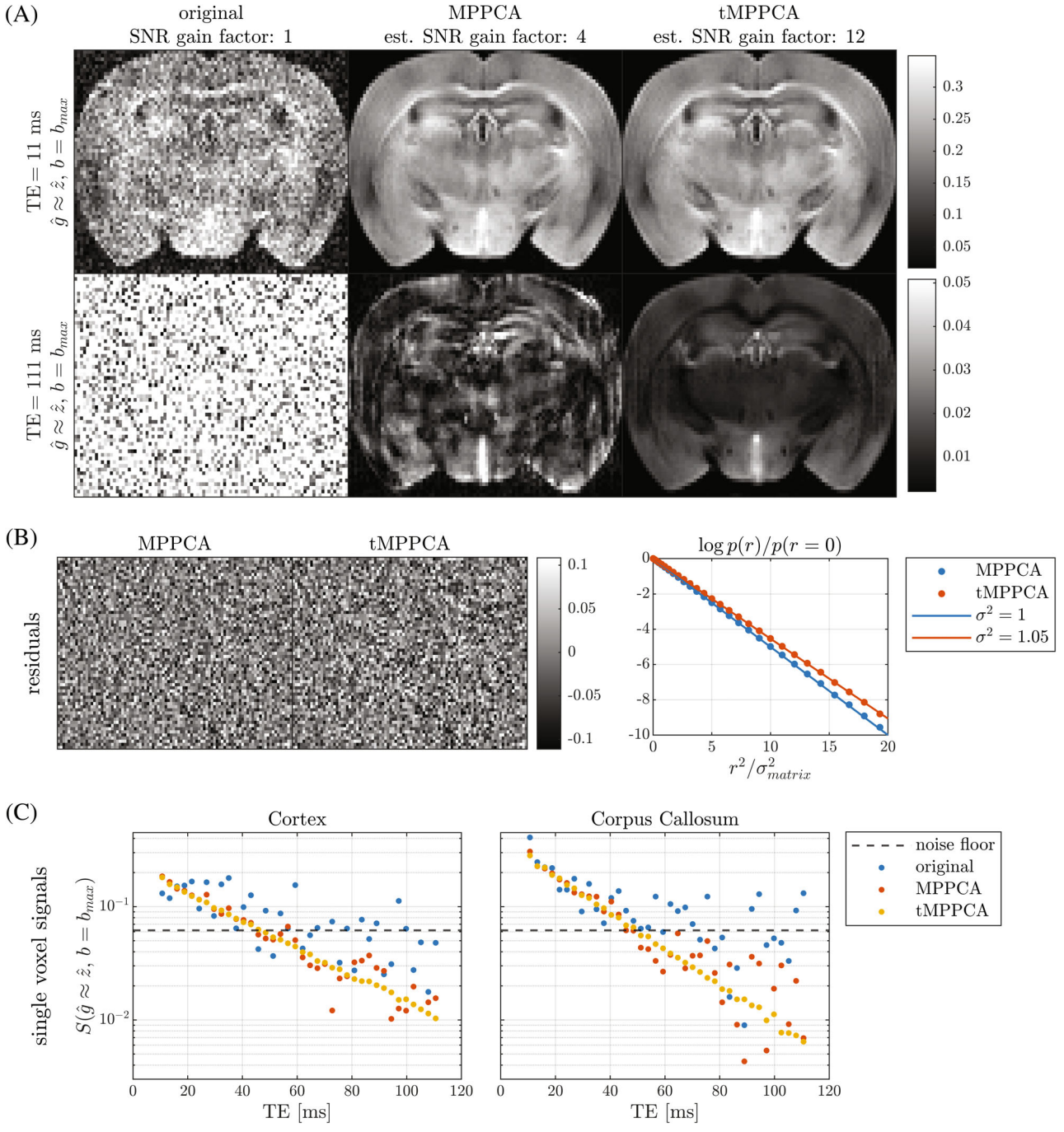

FIGURE 3 Comparison of MP principal component analysis (MPPCA) and tensor MPPCA (tMPPCA) denoising using a 10×10 sliding window and index ordering: voxels $\times$ TE $\times \hat{g} \times b$. A, Magnitude of original and denoised image examples. The color scale is shared for each row. The quality improvement of the recovered images from MPPCA to tMPPCA can be appreciated at large TE. The SNR gains are estimated using Equations (4) and (8), respectively. B, Examples of residual images (real part) and log of the residual distributions compared with Gaussian reference lines. The variance is relative to MPPCA: tMPPCA removes $5\% \approx 1/4^2 - 1/12^2$ additional variance in this data set (Figure 3A). C, Example signals for single voxels located centrally in the cortex and corpus callosum. The remaining noise is not correlated along nonvoxel indices and thus immediately visible. The dashed lines indicate the Rician noise floor for the nondenoised data

window sizes than MPPCA, which peaks beyond the considered window-size interval. Nevertheless, it is implicitly seen that MPPCA also reaches an optimal window size smaller than the image size, because the performance when denoising all voxels as one patch is less than at the otherwise largest considered window. The performance varies slowly with window size around and especially beyond the optimum, suggesting that MPPCA generally is insensitive to the choice of window size so long as it is chosen reasonably close to or larger than the optimal size.

15222594, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mrm.29478 by Cochrane Portugal, Wiley Online Library on [16/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

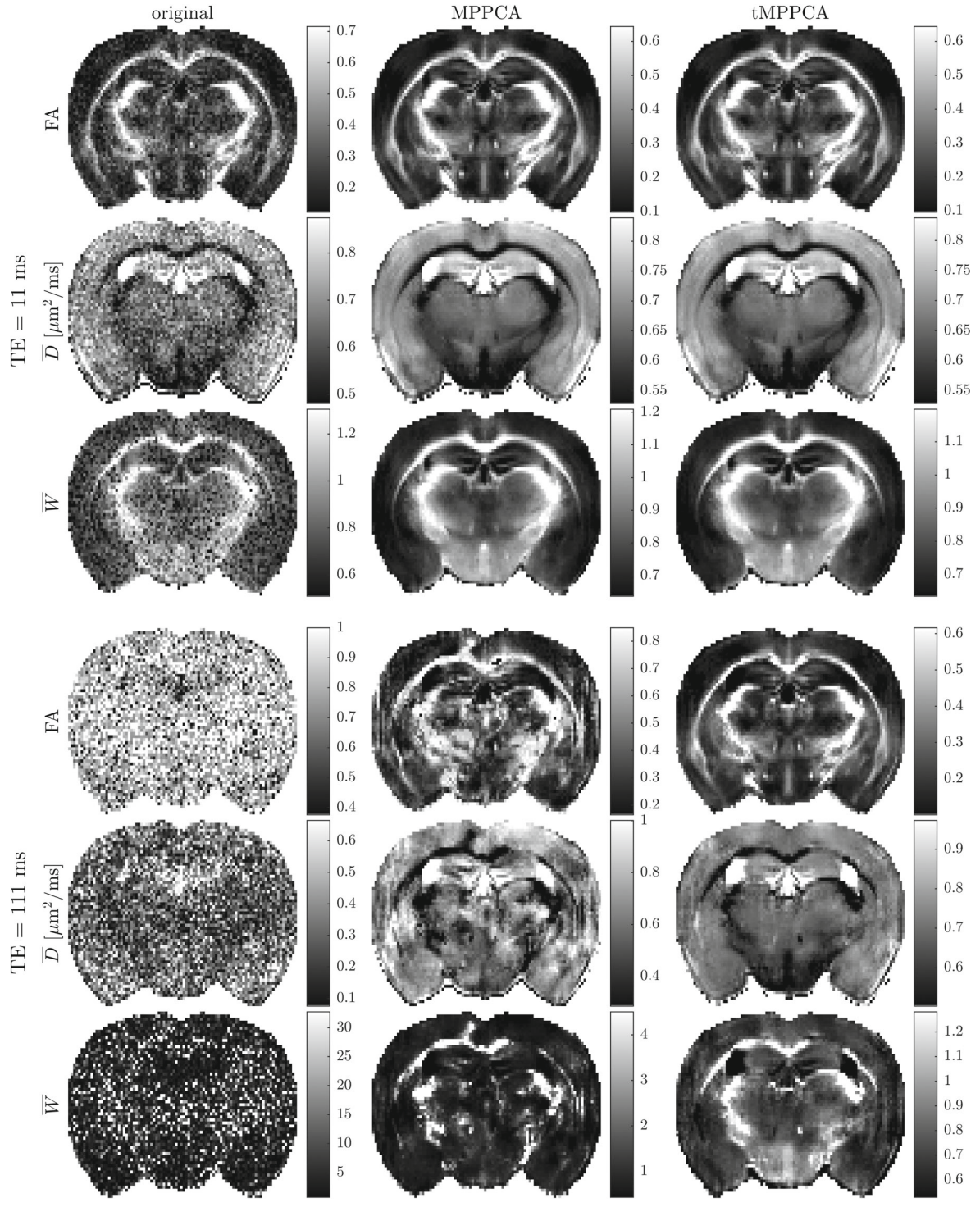

FIGURE 4 Comparison of diffusion parameter maps calculated using the subset of data with smallest/largest TE and denoised with MPPCA and tMPPCA as labeled in the figure. Abbreviation: FA, flip angle

The simulations also indicate that larger window sizes are more beneficial at low SNR and/or for large data sets and vice versa.

Comparing the patch combination methods, we reaffirm that patch averaging provides the best performance with a meaningful performance gap to the method of assigning just the center voxel of each patch. Denoising all voxels as one patch provides competitive performance for large data sets or/and low SNR in the case of MPPCA in accordance with the general dependence on window size. This is not the case for tMPPCA, as it reaches optimal performance at much smaller window size.

15222594, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mrm.29478 by Cochrane Portugal, Wiley Online Library on [16/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

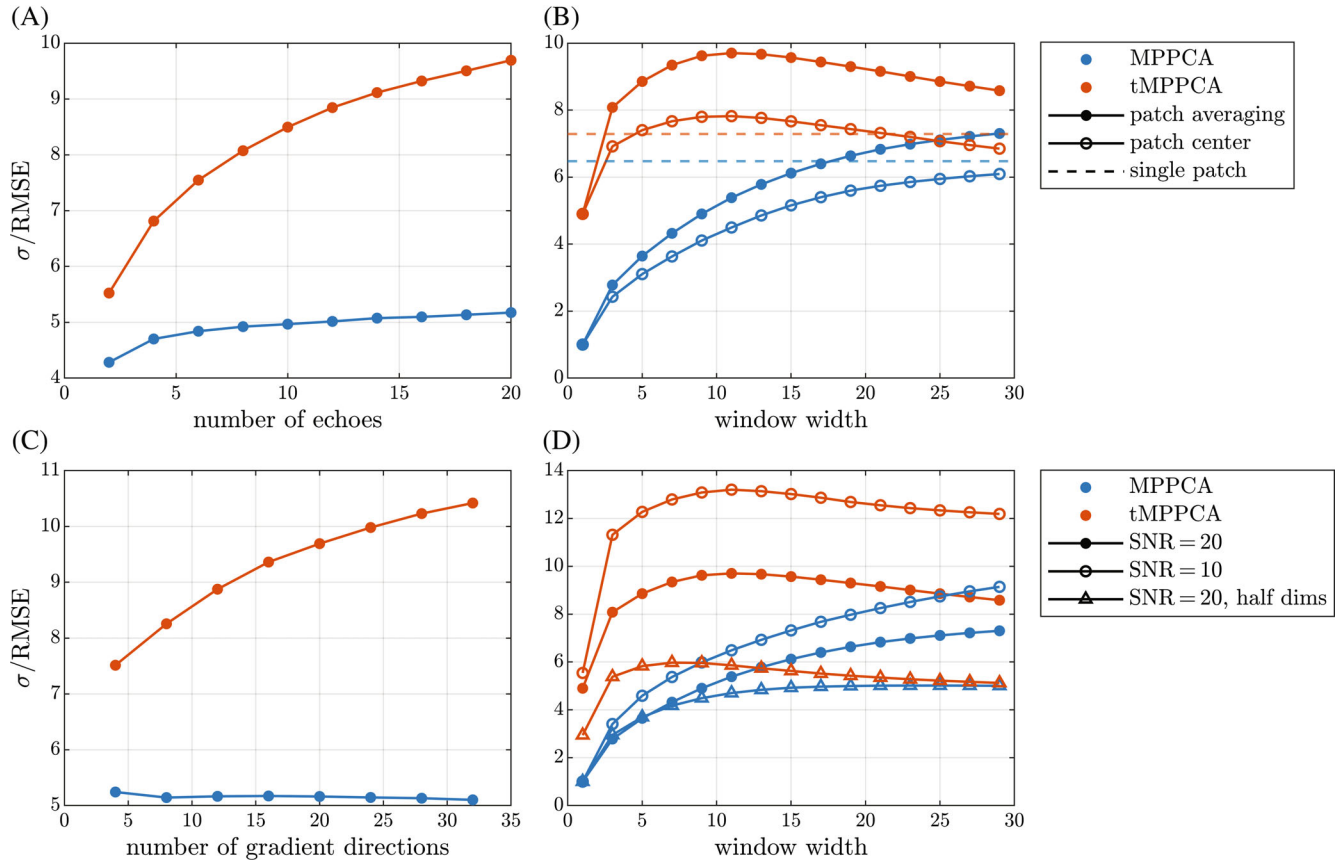

FIGURE 5 The SNR gain from MPPCA and tMPPCA applied to a numerical phantom of multi-TE diffusion data (6 *b*-values from 0.5 to 3.0 μm^2/ms, 20 gradient directions, 20 echoes from 11 to 62 ms, and SNR = 20 unless stated otherwise). A, C, The performance as a function of the number of echoes and gradient directions, respectively, using a 10 × 10 sliding window. B, The performance for varying window size and patch combination methods. Single patch refers to denoising all voxels as one patch. For patch averaging, the denoised signal of each voxel is an average over the contributions from all patches that includes the voxel. For patch center, only the result for the center voxel is used from each denoised patch. D, The performance for varying window size using data sets with different SNR and dimensions. "Half dims" refers to a smaller data set with half number of *b*-values, gradient directions, and echoes. Generally, the index ordering was voxels × TE × $\hat{g}$ × *b* except in the case of single-patch denoising for which it was *TE* × $\hat{g}$ × *b* × voxels. Abbreviation: RMSE, RMS error

5 | DISCUSSION AND CONCLUSIONS

This paper introduced tMPPCA, which generalizes MPPCA[15,16] to exploit the natural tensor structure of multidimensional data. In essence, tMPPCA builds on the success of MPPCA by applying it sequentially when the data structure allows it. Using multi-TE diffusion data as an example, we demonstrated substantial performance improvements in terms of SNR gain using tMPPCA relative to MPPCA. This was strongly indicated by obvious visual improvements of example MR images and signal curves without introducing residuals with spatial structure or deviations from a Gaussian residual distribution. The increased noise suppression improves the potential for dMRI quantification as exemplified by DKI, which showed improvements in terms of quality of the parameter maps. The method can be applied without significantly increased computation time and is in fact faster when compared at peak performance (see subsequently).

The results were verified and extended using a numerical phantom. In particular, we studied the effect of patch size and found that while tMPPCA outperforms MPPCA for any patch size, the performance gap is largest for small sizes. Indeed, tMPPCA was found to reach optimal performance around a 10 × 10 window, whereas MPPCA benefitted from windows beyond 30 × 30. The specific optimal window sizes are dependent on the data, but the results demonstrated that tMPPCA better maintains performance for small patches even with a window size as small as 3 × 3 for the multi-TE diffusion data acquired here. This is presumably an appreciable benefit in scenarios in which smaller windows are preferable due to spatially varying noise.[15,16] Also, in cases with rapidly varying contrast, the trade-off between increasing the window size and thereby increasing the number of signal components increases

15222594, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mrm.29478 by Cochrane Portugal, Wiley Online Library on [16/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the relative advantage of smaller patches. The results also reaffirm the original recommendation of choosing the patch size so that the resulting matrices are approximately square,[15] but this rule of thumb does not apply to tMPPCA.

It is also notable that using smaller patches substantially decreases the computational time required for performing the denoising, which is a considerable benefit in practice. If the initial pass over all indices to improve the σ^2 estimate is skipped, the compute time of tMPPCA will typically be comparable with that of MPPCA: Even though tMPPCA calculates a sequence of SVDs, the dimensionality of the data decreases for each successive SVD, which places most of the computational cost in the initial SVD. If the initial pass is included, the compute time of tMPPCA is longer than that of MPPCA for equal patch sizes. However, if tensor and MPPCA are compared at their respective optimal patch sizes, tMPPCA will be significantly faster because the cost of using larger patches outweighs that of calculating multiple SVDs. For example, MPPCA using a 30×30 window was about 9 times (292 min) slower than tMPPCA (31 min) using a 10×10 window including the additional pass over for an improved σ^2 estimate, which in turn was about 3 times slower than MPPCA (11 min) using a 10×10 window (based on denoising the single slice of the multi-TE dMRI data set with a generic laptop). The specific timings are, however, dependent on hardware and especially the specific data set.

The merits of additional data and redundancy was shown by varying the number of gradient directions and echoes. Both tensor and MPPCA benefits from additional data, but this is much more pronounced for tMPPCA, which better exploits the additional redundancy. This opens several perspectives on particularly favorable applications such as multicoil[20] or multicontrast acquisitions.[35–38] We also note that MRI data are inherently tensor-structured in terms of voxels with x, y, and z indices. For the data considered here, however, we found that combining voxels in one index performs near optimal. Nevertheless, it is conceivable that retaining the voxel tensor structure could increase performance in other scenarios.[30]

We compared the patch combination methods proposed by Veraart et al,[16] namely using only the center voxel of each patch or averaging the contributions from all relevant patches for each voxel as well as denoising all voxels as one patch. Consistent with earlier findings[34] and as exploited in the initial introduction of local PCA to dMRI,[14] patch averaging yields the best performance by a potentially large margin. For large data sets and/or low SNR, the single-patch method can provide comparable performance especially for MPPCA, which benefits more from large window sizes, but presumably also for tMPPCA under specific circumstances. We used the method of equal-weights patch averaging,[15] but tested also weighted averaging with weights chosen according to the estimated residual noise (not shown). Because this did not yield a meaningful performance increase on our test data, we kept to the simpler strategy. However, unequal weighting possibly increases performance in some scenarios.[14,34] Another extension is combining voxels or patches based on another similarity metric than proximity (nonlocal PCA). For example, Zhao et al recently combined MPPCA with a nonlocal strategy involving clustering.[23] Tensor MPPCA could replace MPPCA in such a strategy.

Tensor MPPCA inherits the limitations of MPPCA but does not introduce new ones. These limitations comprise the assumption of independent and identically distributed noise within each patch and the fundamental limitation that it is impossible to fully separate the signal from the noise. Choosing sufficiently small patches can fulfill the requirement of independent and identically distributed noise if the noise level varies spatially. It is also common that the noise level varies in accordance with the signal, as magnitude data are frequently used and are characterized by a Rician or noncentral chi data-noise distribution.[39,40] Sometimes this can be disregarded without a severe performance impact.[15] Alternatively, Zhang et al proposed transforming magnitude dMRI data to equalize the noise variance[41] before HOSVD denoising,[31] and recently Ma et al generalized the approach in a framework combining noise estimation, variance stabilization, and denoising.[24] There, MPPCA with optimal shrinkage was used, but tMPPCA can readily be used with that framework as well. If complex data are available, perhaps the simplest approach is to apply the denoising before calculating the magnitude data as done in the examples here. This also potentially benefits any subsequent reconstruction steps.

That signal and noise cannot be fully separated is partly manifested in inflation of signal singular values, which can be approximately corrected by optimal shrinkage.[33] Furthermore, each signal singular vector is itself randomly rotated proportionally to the noise variance and its inverse signal variance.[33] Accordingly, there will similarly be residual signal variance mixed across the noise components; hence, discarding noise components always entails also removing some true signal variance. Therefore, it would in principle be a misconception to think that MPPCA enables noise suppression without altering the underlying true signal. In scenarios with reasonable SNR, the true signal is almost completely unaffected, but at sufficiently low SNR, MPPCA will remove a significant amount of true signal variance. In practice, it is beneficial for many applications to sacrifice some signal variance to remove a much larger proportion of noise variance, but one can also imagine applications in which retaining the true signal variance has sufficient priority to warrant alternative strategies. For instance, the fitting of robust models

15222594, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mrm.29478 by Cochrane Portugal, Wiley Online Library on [16/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

might be less sensitive to noise than to removal of relatively small amounts of true signal variance in some cases. This is related to the choice of performance metric. Here, we have characterized the denoising performance in terms of least-squared error compared with the true signal. In terms of this metric, it is always beneficial to discard components with variance below the MP upper bound.[33] When applying MPPCA, one should be aware of this limitation even though it will only notably come into play in scenarios with very low SNR, in which case MPPCA may still offer considerable benefit, dependent on the application.

In this paper, we have extended MPPCA to better utilize the redundancy in multidimensional data. Tensor MPPCA yields substantial performance increases for such data sets and performs well for small window sizes, which presumably would be beneficial for instance with spatially varying noise. The generalization introduces no additional assumptions and reduces to MPPCA in the special case of data with rank 2 tensor structure. As such, tMPPCA can be directly adopted as is wherever MPPCA is applied.

ACKNOWLEDGMENT

The authors thank Dmitry Novikov for the discussions. They are also grateful to Mark D. Does and Kevin Harkins from Vanderbilt University for the REMMI pulse sequence and its analysis tools, which were supported through grant number NIH EB019980.

FUNDING INFORMATION

The Danish National Research Foundation and the Danish Ministry of Science, Innovation, and Education (MINDLab); the VELUX Foundation (ARCADIA; 00015963); the European Research Council (679058); "la Caixa" Foundation (100010434); European Union's Horizon 2020 Research and Innovation Program (Marie Skłodowska-Curie grant No. 847648 and fellowship code CF/BQ/PI20/11760029). The authors acknowledge the vivarium of the Champalimaud Center for the Unknown, a facility of CONGENTO that is a research infrastructure co-financed by Lisboa Regional Operational Program (Lisboa 2020), under the PORTUGAL 2020 Partnership Agreement through the European Regional Development Fund (ERDF) and Fundação para a Ciência e Tecnologia (Portugal), project LISBOA-01-0145-FEDER-022170.

DATA AVAILABILITY STATEMENT

Data and matlab code used in the manuscript is freely available at Github (https://github.com/sunenj.

ORCID

Jonas L. Olesen https://orcid.org/0000-0003-4624-9816
Andrada Ianus https://orcid.org/0000-0001-9893-1724
Leif Østergaard https://orcid.org/0000-0003-2930-6997
Noam Shemesh https://orcid.org/0000-0001-6681-5876
Sune N. Jespersen https://orcid.org/0000-0003-3146-4329

REFERENCES

1. Jones DK. *Diffusion MRI Theory, Methods, and Applications*. Oxford University Press; 2010.
2. Novikov DS, Fieremans E, Jespersen SN, Kiselev VG. Quantifying brain microstructure with diffusion MRI: theory and parameter estimation. *NMR Biomed*. 2019;32:e3998.
3. Jones DK. Precision and accuracy in diffusion tensor magnetic resonance imaging. *Top Magn Reson Imaging*. 2010;21:87-99.
4. Lattanzi R, Sodickson DK. Ideal current patterns yielding optimal signal-to-noise ratio and specific absorption rate in magnetic resonance imaging: computational methods and physical insights. *Magn Reson Med*. 2012;68:286-304.
5. Fan Q, Witzel T, Nummenmaa A, et al. MGH–USC human connectome project datasets with ultra-high b-value diffusion MRI. *Neuroimage*. 2016;124:1108-1114.
6. Ocali O, Atalar E. Ultimate intrinsic signal-to-noise ratio in MRI. *Magn Reson Med*. 1998;39:462-473.
7. Baltes C, Radzwill N, Bosshard S, Marek D, Rudin M. Micro MRI of the mouse brain using a novel 400 MHz cryogenic quadrature RF probe. *NMR Biomed*. 2009;22:834-842.
8. Chen G, Wu Y, Shen D, Yap PT. XQ-NLM: denoising diffusion MRI data via x-q space non-local patch matching. *Med Image Comput Comput Assist Interv*. 2016;9902:587-595.
9. Coupé P, Yger P, Prima S, Hellier P, Kervrann C, Barillot C. An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. *IEEE Trans Med Imaging*. 2008;27:425-441.
10. Manjón JV, Coupé P, Buades A, Louis Collins D, Robles M. New methods for MRI denoising based on sparseness and self-similarity. *Med Image Anal*. 2012;16:18-27.
11. Knoll F, Bredies K, Pock T, Stollberger R. Second order total generalized variation (TGV) for MRI. *Magn Reson Med*. 2011;65:480-491.
12. Muckley MJ, Ades-Aron B, Papaioannou A, et al. Training a neural network for Gibbs and noise removal in diffusion MRI. *Magn Reson Med*. 2021;85:413-428.
13. Tian Q, Li Z, Fan Q, et al. SDnDTI: self-supervised deep learning-based denoising for diffusion tensor MRI. *Neuroimage*. 2022;253:119033.
14. Manjón JV, Coupé P, Concha L, Buades A, Collins DL, Robles M. Diffusion weighted image denoising using overcomplete local PCA. *PLoS One*. 2013;8:e73021.
15. Veraart J, Novikov DS, Christiaens D, Ades-aron B, Sijbers J, Fieremans E. Denoising of diffusion MRI using random matrix theory. *Neuroimage*. 2016;142:394-406.
16. Veraart J, Fieremans E, Novikov DS. Diffusion MRI noise mapping using random matrix theory. *Magn Reson Med*. 2016;76:1582-1593.
17. Marčenko VA, Pastur LA. Distribution of eigenvalues for some sets of random matrices. *Math USSR-Sbornik*. 1967;1:457-483.
18. Ding Y, Chung YC, Simonetti OP. A method to assess spatially variant noise in dynamic MR image series. *Magn Reson Med*. 2010;63:782-789.

15222594, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mrm.29478 by Cochrane Portugal, Wiley Online Library on [16/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19. Sengupta AM, Mitra PP. Distributions of singular values for some random matrices. *Phys Rev E*. 1999;60:3389-3392.
20. Lemberskiy G, Baete S, Veraart J, Shepherd TM, Fieremans E, Novikov DS. Achieving sub-mm clinical diffusion MRI resolution by removing noise during reconstruction using random matrix theory. *In: Proceedings of the 27th Annual Meeting of ISMRM, Montréal, Canada*. 2019;27:0770.
21. Cordero-Grande L, Christiaens D, Hutter J, Price AN, Hajnal JV. Complex diffusion-weighted image estimation via matrix recovery under general noise models. *Neuroimage*. 2019;200:391-404.
22. Does MD, Olesen JL, Harkins KD, et al. Evaluation of principal component analysis image denoising on multi-exponential MRI relaxometry. *Magn Reson Med*. 2019;81:3503-3514.
23. Zhao W, Lv Y, Liu Q, Qin B. Detail-preserving image denoising via adaptive clustering and progressive PCA thresholding. *IEEE Access*. 2018;6:6303-6315.
24. Ma X, Uğurbil K, Wu X. Denoise magnitude diffusion magnetic resonance images via variance-stabilizing transformation and optimal singular-value manipulation. *Neuroimage*. 2020;215:116852.
25. Ades-Aron B, Lemberskiy G, Veraart J, et al. Improved task-based functional MRI language mapping in patients with brain tumors through marchenko-pastur principal component analysis denoising. *Radiology*. 2020;298:365-373.
26. Ianuş A, Carvalho J, Fernandes FF, et al. Soma and Neurite density MRI (SANDI) of the in-vivo mouse brain and comparison with the Allen Brain Atlas. *NeuroImage*. 2022;254:119135.
27. Kim Y, Chen HY, Autry AW, et al. Denoising of hyperpolarized 13C MR images of the human brain using patch-based higher-order singular value decomposition. *Magn Reson Med*. 2021;86:2497-2511.
28. Wang L, Xiao D, Hou WS, Wu XY, Chen L. A modified higher-order singular value decomposition framework with adaptive multilinear tensor rank approximation for three-dimensional magnetic resonance Rician noise removal. *Front Oncol*. 2020;10:1640.
29. Chen HY, Autry AW, Brender JR, et al. Tensor image enhancement and optimal multichannel receiver combination analyses for human hyperpolarized 13C MRSI. *Magn Reson Med*. 2020;84:3351-3365.
30. Brender JR, Kishimoto S, Merkle H, et al. Dynamic imaging of glucose and lactate metabolism by 13 C-MRS without hyperpolarization. *Sci Rep*. 2019;9:1-14.
31. Zhang X, Peng J, Xu M, et al. Denoise diffusion-weighted images using higher-order singular value decomposition. *Neuroimage*. 2017;156:128-145.
32. Zhang X, Xu Z, Jia N, et al. Denoising of 3D magnetic resonance images by using higher-order singular value decomposition. *Med Image Anal*. 2015;19:75-86.
33. Gavish M, Donoho DL. Optimal shrinkage of singular values. *IEEE Trans Inf Theory*. 2017;63:2137-2152.
34. Katkovnik V, Foi A, Egiazarian K, Astola J. From local kernel to nonlocal multiple-model image denoising. *Int J Comput Vis*. 2009;86:1-32.
35. Veraart J, Novikov DS, Fieremans E. TE dependent diffusion imaging (TEdDI) distinguishes between compartmental T2 relaxation times. *Neuroimage*. 2018;182:360-369.
36. Lampinen B, Szczepankiewicz F, Mårtensson J, et al. Towards unconstrained compartment modeling in white matter using diffusion-relaxation MRI with tensor-valued diffusion encoding. *Magn Reson Med*. 2020;84:1605-1623.
37. Kleban E, Tax CMW, Rudrapatna US, Jones DK, Bowtell R. Strong diffusion gradients allow the separation of intra- and extra-axonal gradient-echo signals in the human brain. *Neuroimage*. 2020;217:116793.
38. Shemesh N, Jespersen SN, Alexander DC, et al. Conventions and nomenclature for double diffusion encoding NMR and MRI. *Magn Reson Med*. 2016;75:82-87.
39. Gudbjartsson H, Patz S. The rician distribution of noisy mri data. *Magn Reson Med*. 1995;34:910-914.
40. Coupé P, Manjón JV, Gedamu E, Arnold D, Robles M, Collins DL. Robust Rician noise estimation for MR images. *Med Image Anal*. 2010;14:483-493.
41. Foi A. Noise estimation and removal in MR imaging: the variance-stabilization approach. *In: Proceedings of the International Symposium on Biomedical Imaging, Chicago, Illinois, USA*. 2011;1809-1814.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

FIGURE S1. Comparison of MPPCA and tMPPCA denoising using a single patch including all voxels and index-ordering: voxels $\times (b, \hat{g}) \times$ TE. (A) Magnitude of original and denoised image examples. The SNR gains are estimated using Equation (4) and (8) respectively. (B) Examples of residual images (real part) and log of the residual distributions compared to Gaussian reference lines. (C) Example signals for single voxels located centrally in the Cortex and Corpus Callosum

FIGURE S2. Comparison of MPPCA and tMPPCA denoising using a 3×3 sliding window and index-ordering TE $\times$ $\hat{g}$ $\times$ voxels $\times$ b. (A) Magnitude of original and denoised image examples. The SNR gains are estimated using Equation (4) and (8) respectively. (B) Examples of residual images (real part) and log of the residual distributions compared to Gaussian reference lines. (C) Example signals for single voxels located centrally in the Cortex and Corpus Callosum

How to cite this article: Olesen JL, Ianus A, Østergaard L, Shemesh N, Jespersen SN. Tensor denoising of multidimensional MRI data. *Magn Reson Med*. 2023;89:1160-1172. doi: 10.1002/mrm.29478

15222594, 2023, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mrm.29478 by Cochrane Portugal, Wiley Online Library on [16/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License