IEEE Open Journal of Signal Processing

Received 29 March 2021; revised 16 July 2021; accepted 16 July 2021. Date of publication 29 July 2021;
date of current version 3 September 2021. The review of this article was arranged by Associate Editor A. Majumdar.

Digital Object Identifier 10.1109/OJSP.2021.3100751

Dictionary+Wavelet Model With Nested-Minorized VB-EM for SMS-CAIPI R-fMRI Reconstruction

PRACHI H. KULKARNI [1], S. N. MERCHANT [1], AND SUYASH P. AWATE [2]

[1] Department of Electrical Engineering, IIT Bombay, Mumbai 400076, India
[2] Department of Computer Science and Engineering, IIT Bombay, Mumbai 400076, India

CORRESPONDING AUTHOR: PRACHI H. KULKARNI (e-mail: phkulkarni@ee.iitb.ac.in)

This work was supported by the Infrastructure Facility for Advanced Research and Education in Diagnostics, Department of Biotechnology, Government of India under Grant BT/INF/22/SP23026/2017.

ABSTRACT Resting-state functional magnetic resonance imaging (R-fMRI) applications can entail a higher temporal-sampling rate that trades off spatial resolution, thereby challenging effective scientific studies. To enable higher spatial resolution, current schemes speedup per-timeframe scanning by reconstruction from simultaneous multislice (SMS) magnetic resonance imaging (MRI) with k-space undersampling (sometimes temporal undersampling), while using prior models on the signal. We propose a novel *algorithmic framework* to reconstruct R-fMRI (SMS with controlled aliasing) that has, *both*, k-space undersampling and temporal undersampling. We propose a *coupled* spatiotemporal sparse model, incorporating (i) a robust spatially-regularized temporal-dictionary prior and (ii) a spatiotemporal wavelet prior, which we fit efficiently using *variational Bayesian expectation maximization with nested minorization* (VBEMNM). We show that our framework has the potential to enable higher spatial resolution without increasing scan time in R-fMRI that has inherently weak signals and is therefore prone to large physiological fluctuations, acquisition noise, and imaging artifacts. Qualitative and quantitative evaluation on retrospectively undersampled brain R-fMRI shows that estimates of resting-state networks from our framework and the boost in temporal stability given by our framework compares favourably to existing methods for R-fMRI reconstruction.

INDEX TERMS R-fMRI, reconstruction, SMS with CAIPI, joint k-t undersampling, coupled dictionary and wavelets, robust, variational Bayesian EM, nested minorization.

I. INTRODUCTION

Functional magnetic resonance imaging (fMRI) acquires a 4D spatiotemporal image to assess the functional areas in the cerebral cortex, unlike structural MRI that acquires a 3D spatial volume to assess structural aspects, e.g., injuries or tumors. Brain fMRI acquisition is performed as a collection of transaxial slices, stackable as a spatial volume, at each timeframe. To speedup acquisition, multiple transaxial slices can be acquired simultaneously as a linear combination. This is known as simultaneous multislice (SMS) imaging. In this case, the fMRI scanner acquires the undersampled Fourier-transform of the linear combination of multiple transaxial slices. This imaging in the frequency domain is also known as imaging in k-space. We describe the explicit forward model for fMRI acquisition later in Section III.

In task-based fMRI, at voxels within the brain areas activated by the task, the blood oxygenation level dependent (BOLD) signals exhibit temporal correlations. Thus, task-based fMRI aids in determining whether the task-appropriate functional area is activated. In some cases, task-based fMRI faces limitations when patients are unable to cooperate with the prescribed tasks because of age (young children, elderly, etc.) or because of some neuropsychiatric disorder. Also, a task-based fMRI scan highlights only the task-specific functional area of the brain and may be unable to assess the overall functional health of the brain. In contrast, resting-state fMRI (R-fMRI) [1] can overcome some of these challenges by imaging the subject during a wakeful resting state or during sedation. R-fMRI leverages the fact that, within voxels in functionally connected areas or resting-state networks (RSN),

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

the BOLD time-series exhibit correlations even when the subject is in resting state [1]. Such correlations are consistent across subjects within a homogeneous cohort, despite significant variability across subjects' BOLD time-series [1], [2].

There are certain practical limitations that hamper the utility of R-fMRI. The BOLD signal underlying R-fMRI comprises low frequencies that are typically less than 0.1 Hz. Further, because the imaging is in resting state, the R-fMRI signal is typically weak and thereby prone to degradations in the form of large physiological fluctuations, acquisition noise, and imaging artifacts leading to outliers [3], [4]. Reducing the effects of the degradations requires a trade-off between temporal sampling rates and spatial resolution. Higher temporal sampling rates limit the acquisition duration for each timeframe and, thereby, the spatial resolution. Higher spatial resolution limits the temporal sampling rate risking aliasing from the high-frequency components in the degradations. Typical R-fMRI employs temporal sampling rates that are about an order of magnitude higher than 0.1 Hz [2], limiting the spatial resolution to 8–64 mm^3 voxels and affecting the studies of the cerebral cortex that is typically 3–4 mm thick.

Typical applications of R-fMRI can benefit greatly from higher spatiotemporal resolution through improved signal sensitivity and scientific-study reliability. However, current commercial imaging systems need a trade-off between high temporal resolution and high spatial resolution. The need to reduce this trade-off motivates the speeding up of the per-timeframe R-fMRI acquisition, which currently uses reconstruction methods relying on (i) SMS imaging with k-space undersampling [5], [6] or (ii) only k-space (and sometimes time) undersampling with priors on the signal [7]–[9].

This paper makes novel contributions. We propose an *algorithmic framework* to reconstruct *SMS* R-fMRI (with controlled aliasing) that is undersampled in *both* k-space and time. This can potentially enable a $3\times$ improvement in resolution along each voxel dimension (i.e., $27\times$ reduced voxel volume), without increasing the total scan time. The framework alleviates the effects of large degradations by leveraging a robust sparse model that *couples* a *spatially-regularized temporal dictionary* model with a spatiotemporal *wavelet* model. We propose a reconstruction method using variational Bayesian expectation maximization with nested minorization (VBEMNM) that leads to *efficient* parameter updates within each iteration. Thus, we refer to our framework as DW+VBEMNM. Qualitative and quantitative evaluation on retrospectively undersampled brain R-fMRI shows that estimates of resting-state networks from our framework and the boost in temporal stability given by our framework compares favourably to existing methods for R-fMRI reconstruction.

II. RELATED WORK

R-fMRI image analysis for general clinical applications [2] uses many methods including seed-based analysis, independent component analysis, graph-based analysis [2], and deep learning [10]. To reconstruct undersampled R-fMRI, one class of methods uses complex pulse sequences like multiplexed echo planar imaging, SMS imaging, or parallel imaging using multiple coils coupled with k-space undersampling [5], [6]. Reconstructing parallel MRI typically leverages variants/combinations of two very popular techniques: sensitivity encoding (SENSE) [11] and generalized auto-calibrating partially parallel acquisitions (GRAPPA) [12]. SENSE solves the linear inverse problem in SMS imaging using estimates of coil-sensitivity profiles. GRAPPA first learns a linear kernel from reference k-space data, i.e., the auto-calibration signal (ACS), to fill-in missing k-space lines, for each coil, then uses per-coil inverse Fourier transforms with root-sum-of-squares scheme [13] to give the reconstructed image. SMS MRI reduces scan time by a factor equal to the number of slices imaged simultaneously. To reduce the ill-posed nature of the inverse problem in the reconstruction of SMS MRI, slice-specific spatial shifts are implicitly introduced through explicit slice-specific phase shifts in k-space. This is known as controlled aliasing in parallel imaging (CAIPI) [14]. Other methods based on SENSE [15], sometimes using CAIPI [16], and GRAPPA [17], [18] or a combination of both are also used. Another class of methods speedup R-fMRI without using SMS imaging, but by reconstructing k-space undersampled (and sometimes temporally-undersampled) data using priors on the R-fMRI signal, e.g., low-rank models in the k-space $\times$ time space [8], [19], [20], sparse models in wavelet domain [9], or dictionary-based models [21]. In contrast, our framework can speedup R-fMRI acquisition by a larger factor by combining SMS imaging (with CAIPI) with undersampling in both k-space and time using a robust sparse spatially-regularized coupled dictionary+wavelet model.

Non-Cartesian k-space sampling trajectories [19], [20], [22]–[25] can enable higher undersampling factors but might lead to artifacts and might be non-trivial to implement on typical commercial scanners. In contrast, we employ Cartesian undersampling in k-space that is easily implementable on typical commercial scanners. Some methods employ matrix recovery [26], [27] and blind compressive sensing [28] for general dynamic MRI, but they undersample only in k-space. Undersampling jointly in the k-space domain as well as the temporal domain (involving a smaller temporal sampling rate) can enable higher undersampling factors. In such scenarios of temporal undersampling, with missing timeframes, it can be non-trivial to extend approaches based on matrix recovery or blind compressive sensing.

In medical image analysis, estimating per-voxel uncertainty/confidence in reconstructed images can be useful to determine the reliability of the reconstructions and in subsequent image analysis [29], [30]. However, attention to uncertainty estimation has been limited in medical image reconstruction [31]–[34], in general, and in R-fMRI reconstruction [29]. We estimate the per-voxel uncertainty in reconstructed images to potentially identify regions that were difficult to reconstruct reliably, for many possible reasons.

Some of the basic methods used for cleaning the fMRI signal during preprocessing include nuisance regression [35] and ICA-based methods like FIX [36]. However, the preprocessed

signal may still contain large physiological fluctuations which can be reduced using standard denoising methods like dictionary learning and sparse coding (DLSC) [37] or temporal non-local means (tNLM) [38]. A recent method uses deep learning [39] for denoising task-based fMRI, but is not applicable to R-fMRI. In contrast, we use a robust sparse spatially-regularized dictionary model coupled with a spatiotemporal wavelet model for reconstruction with denoising capability.

This paper significantly extends our preliminary work on k-t undersampled SMS-CAIPI R-fMRI reconstruction in [7]. While [7] proposes k-t undersampled SMS-CAIPI R-fMRI using a subject-invariant spatially-regularized dictionary prior model, this paper proposes several improvements over [7]. First, while [7] used the dictionary model alone, our framework in this paper couples the dictionary model with a 3D spatiotemporal wavelet model for additional spatiotemporal regularization. Second, while [7] used simple projected gradient ascent with adaptive learning rate for optimization, our framework in this paper uses the more sophisticated VBEM-MNM [40] optimization scheme. We show that both these improvements result in better RSNs than those in [7]. Third, our framework in this paper is able to generate per-voxel maps showing the uncertainty of reconstruction, unlike [7]. Fourth, while [7] shows results for only two RSNs, this paper shows results for four RSNs. Fifth, this paper shows empirical evaluation on a larger cohort.

III. METHODS

We describe our novel framework for reconstructing R-fMRI from SMS-CAIPI acquisitions with joint k-t undersampling. We describe our robust image model coupling a spatially-regularized temporal dictionary and a spatiotemporal wavelet frame. We describe our optimization algorithm to handle the non-concavity of our robust model using nested minorization within VBEM, giving efficient parameter updates within each iteration of EM.

A. COUPLED ROBUST DICTIONARY+WAVELET PRIOR MODEL

Notation. Let the Markov random field (MRF) X model the 4D spatiotemporal complex-valued R-fMRI BOLD-signal image having L slices, V voxels in each slice and T time-frames. Let $X_{(l,v)} \in \mathbb{C}^T$ be the time-series at voxel (l, v). Let $X^t \in \mathbb{C}^{LV}$ be the 3D spatial image at timeframe t. Let $\mathcal{N}^X := \{\mathcal{N}^X_{(t,(l,v))} : l \in [1, L], v \in [1, V], t \in [1, T]\}$ be a neighborhood system, where, for voxel (l, v) at time t, neighborhood $\mathcal{N}^X_{(t,(l,v))} := \{(t', (\boldsymbol{l'}, \boldsymbol{v'})) : t' \neq t, l' = l, v' = v\}$ includes all timeframes $(t', (l, v))$ in the time-series at voxel (l, v), excluding $(t, (l, v))$ itself. Let $\overline{X_{(l,v)}} \in \mathbb{R}^T_{\geq 0}$ be the magnitude of BOLD-signal time-series $X_{(l,v)}$ at voxel (l, v).

Dictionary+Wavelet Model. We propose a robust model for the BOLD signal using a dictionary D with J unit-norm atoms for the per-voxel time-series $X_{(l,v)}$ (j-th atom $d_j \in \mathbb{R}^T \equiv j$-th column of D), coupled with a 2D+t spatiotemporal dual-tree-complex (overcomplete) wavelet [41] frame W. Our robust dictionary model provides a strong temporal prior on the R-fMRI signal, along with some spatial regularization. The dictionary model is learned from a limited amount of R-fMRI data that is known to have lower signal-to-noise ratio (SNR) than task-based fMRI. On the other hand, the wavelet model provides a rich analytically-designed spatiotemporal model with nice theoretical properties. Thus, we propose a coupled modeling approach to leverage the benefits of both models.

Dictionary Model. Associated with the dictionary D, let A be the coefficient matrix where the the column corresponding to the voxel (l, v) is the coefficient vector $a_{(l,v)} \in \mathbb{R}^J$. Let $\mathcal{N}^A_{(l,v)}$ comprise the 26 neighbours of the voxel (l, v) in a $3\times3\times3$ neighborhood. We model, as detailed in [7], the time-series $X_{(l,v)}$ at voxel (l, v) as $\Phi_{(l,v)}(RDa_{(l,v)} + d_0 a_{0(l,v)})$, where (i) $\Phi_{(l,v)}$ is a $T \times T$ complex-valued diagonal matrix that incorporates the complex phase of the time-series, (ii) the coefficients $a_{0(l,v)}$ and $a_{(l,v)}$, respectively, correspond to the constant atom $d_0 := \mathbf{1}$ and the other J atoms in D, and (iii) R is a $T \times T$ rotation-transform matrix.

RSNs characterized by high normalized-cross-correlations within networks are quite consistent across the population of healthy subjects in spite of the inter-session and inter-subject variability in the underlying time-series $X_{(l,v)}$. This consistency in normalized-cross-correlations is modeled using the similarity transform on the time-series where the scaling is handled by the dictionary coefficients $a_{(l,v)}$, shifts by the mean value $d_0 a_{0(l,v)}$, and the rotation by R. To motivate this further, we observe that, if we transform two time-series, modeled as vectors, using two associated similarity transforms involving the same rotation, then the normalized cross-correlation between the original time-series is the same as that between the transformed time-series. Thus, our proposed use of the similarity transform adapts the dictionary to an arbitrary session/subject (to give good temporal-signal fits) while maintaining the consistency of the normalized-cross-correlation maps across sessions and subjects.

Our framework models heavy-tailed distributions [42] on the residual $(X_{(l,v)} - \Phi_{(l,v)}(RDa_{(l,v)} + d_0 a_{0(l,v)}))$, at all voxels (l, v). The prior model [7] on the dictionary coefficients A, enforcing sparsity to prevent overfitting and spatial regularization for smooth discontinuity-adaptive RSNs, is

$$P(A) \propto \exp\Bigg(- \sum_{l=1}^{L}\sum_{v=1}^{V} \alpha \|a_{(l,v)}\|^q_{q,\epsilon} - \sum_{l=1}^{L}\sum_{v=1}^{V}\sum_{j=1}^{J} \sum_{(\boldsymbol{l'},\boldsymbol{v'})\in\mathcal{N}^A_{(l,v)}} \beta\omega_{(l,v),(\boldsymbol{l'},\boldsymbol{v'})} H_\gamma(a^j_{(l,v)} - a^j_{(\boldsymbol{l'},\boldsymbol{v'})})\Bigg), \tag{1}$$

where we define $\|z\|^q_{q,\epsilon} := \sum_i (z_i^2 + \epsilon)^{q/2}$; $\alpha \in \mathbb{R}^+, \beta \in \mathbb{R}^+, \gamma \in \mathbb{R}^+$, $q \in (0, 1)$ are free parameters; q controls the sparsity on the dictionary fits; the Huber loss function $H_\gamma(a^j_{(l,v)} - a^j_{(\boldsymbol{l'},\boldsymbol{v'})})$ equals $0.5(a^j_{(l,v)} - a^j_{(\boldsymbol{l'},\boldsymbol{v'},v')})^2$ when

$|a^j_{(l,v)} - a^j_{(\boldsymbol{l'},\boldsymbol{v'})}| \leqslant \gamma$, and equals $\gamma|a^j_{(l,v)} - a^j_{(\boldsymbol{l'},\boldsymbol{v'})}| - 0.5\gamma^2$ when $|a^j_{(l,v)} - a^j_{(\boldsymbol{l'},\boldsymbol{v'})}| > \gamma$; and the weights $\omega_{(l,v),(l',v')}$ model a Gaussian decay with increasing spatial distance between the voxels (l, v) and $(\boldsymbol{l'}, \boldsymbol{v'})$. While the quasi-norm enforces sparsity on the dictionary coefficients, the Huber loss enforces discontinuity-adaptive spatial regularization by enforcing sparsity on the differences between the dictionary coefficients within spatial neighborhoods.

Wavelet Model. Let F be the coefficient matrix for the wavelet frame W, where the the column corresponding to the slice l is the coefficient vector f_l. We model the signal X_l at slice l as Wf_l. The wavelet model employs a Gaussian distribution for the residual $X_l - Wf_l$ for all slices l. The prior model on the wavelet coefficients F that enforces sparsity to prevent overfitting, is modeled as

$$P(F) \propto \exp\Big(-\sum_{l=1}^{L} \rho \|f_l\|^1_{1,\epsilon}\Big), \tag{2}$$

where $\rho \in \mathbb{R}^+$ is a free parameter.

Coupled Model. Combining the dictionary and wavelet models leads to the prior on X as

$$P(X|R, A, \Phi, F; D, W) \propto \exp\Bigg(-\sum_{l=1}^{L}\sum_{v=1}^{V} \delta\|X_{(l,v)} - \Phi_{(l,v)}(RDa_{(l,v)} + d_0 a_{0(l,v)})\|^p_{2,\epsilon} - \sum_{l=1}^{L} \eta\|X_l - Wf_l\|^2_{2,\epsilon}\Bigg), \tag{3}$$

where we define $\|z\|^p_{2,\epsilon} := (\|z\|^2_2 + \epsilon)^{p/2}$ and δ, η and $p \in (0, 2)$ are free parameters. p controls the heaviness of the tail of the robust dictionary model. Thus, we propose the generative dictionary+wavelet prior model on the R-fMRI image as $P(X|R, A, \Phi, F; D, W)P(A)P(F)$.

Dictionary Learning. We learn the dictionary D from the temporally-filtered (to retain only the R-fMRI frequencies less than 0.1 Hz) R-fMRI data X of just one subject, by maximizing the generative dictionary+wavelet prior $P(X|R, A, \Phi, F; D, W)P(A)P(F)$ with respect to D and A, where R is fixed to the $T \times T$ identity matrix and $\Phi_{(l,v)}$ are fixed to the phase values of $X_{(l,v)}$. We initialize (i) the atoms d_j to the centroids obtained through k-means on all $\overline{X_{(l,v)}}$ which are further normalized to have zero mean and unit norm, and (ii) the coefficients $a^j_{(l,v)}$ to the fractions $a^j_{(l,v)} := \|\overline{X_{(l,v)}}\|_2 \exp(-\sigma_A\|(\overline{X_{(l,v)}}/\|\overline{X_{(l,v)}}\|_2) - d_j\|^2_2)/\max_k \exp(-\sigma_A\|(\overline{X_{(l,v)}}/\|\overline{X_{(l,v)}}\|_2) - d_k\|^2_2)$, where $\sigma_A \in \mathbb{R}^+$ is a free parameter. We maximize the log-prior using projected gradient ascent with adaptive learning rate, with the dictionary atoms d_j projected onto the subspace of unit-norm atoms after each update. We find that the data of only one subject is sufficient to learn the dictionary, because the similarity transform enables the dictionary D to adapt to different subjects while maintaining the consistency in RSNs.

B. LIKELIHOOD MODEL

Let K be the number of coils, where Ψ^k_l is the coil sensitivity profile for coil k and slice l. Let M consecutive slices be imaged simultaneously for SMS with CAIPI [14], and let $\mathcal{C}_l$ model the implicit spatial shift underlying CAIPI. Let N be the number of sets of M simultaneously-imaged slices in whole-brain imaging. Let Y be the k-space representation of the underlying R-fMRI image X, where $Y^{(k,t)}_n$ is the k-space representation of X corresponding to coil k for slice-set n at time t. $Y^{(k,t)}_n$ comprises (i) the observed/acquired k-t undersampled SMS-CAIPI data $Y^{(k,t)}_{\mathcal{O},n}$ and (ii) the missing k-space values $Y^{(k,t)}_{\mathcal{M},n}$. For each coil, we model the acquisition noise in k-space as independent and identically distributed (i.i.d.) additive complex Gaussian, having zero mean and variance σ^2 in the real and imaginary parts. We estimate σ^2 by computing the variance in the background of a fully-sampled R-fMRI acquisition. Let $\mathcal{F}$ be the Fourier transform operator. Let $\mathcal{S}^t_{\mathcal{O},n}$ and $\mathcal{S}^t_{\mathcal{M},n}$ represent sampling operators for slice-set n at time t that select the k-space values at the observed and missing frequencies, respectively. $\mathcal{S}^t_{\mathcal{O},n}$ and $\mathcal{S}^t_{\mathcal{M},n}$ also model temporal undersampling with $\mathcal{S}^t_{\mathcal{O},n}$ selecting a null set of k-space values and $\mathcal{S}^t_{\mathcal{M},n}$ selecting all the k-space values for the missing time-frames. Thus, the forward model for observed/acquired k-t undersampled SMS-CAIPI R-fMRI data for coil k, slice-set n at time t is

$$Y^{(k,t)}_{\mathcal{O},n} = \mathcal{S}^t_{\mathcal{O},n}\mathcal{F}\sum_{l=1+(n-1)M}^{nM} \mathcal{C}_l\Psi^k_l X^t_l + \xi, \tag{4}$$

where ξ is i.i.d. complex-Gaussian noise, having zero mean and variance σ^2 in the real and imaginary parts.

Thus, the likelihood of the observed k-t undersampled SMS-CAIPI k-space data $Y_{\mathcal{O}}$ is

$$P(Y_{\mathcal{O}}|X) \propto \exp\Bigg(-\sum_{k=1}^{K}\sum_{n=1}^{N}\sum_{t=1}^{T}(0.5/\sigma^2) \|Y^{(k,t)}_{\mathcal{O},n} - \mathcal{S}^t_{\mathcal{O},n}\mathcal{F}\sum_{l=1+(n-1)M}^{nM} \mathcal{C}_l\Psi^k_l X^t_l\|^2_2\Bigg). \tag{5}$$

C. EXPECTATION MAXIMIZATION (EM) FRAMEWORK FOR INFERENCE

Given the observed k-space data $Y_{\mathcal{O}}$, we employ the generalized EM framework for inference. We choose X as the latent random variable and $\theta := \{Y_{\mathcal{M}}, R, A, \Phi, F\}$ as the set of parameters. EM aims to maximize the incomplete-data log-posterior $\log(\zeta P(Y_{\mathcal{O}}, \theta))$, with ζ as the normalizing constant. The incomplete-data log-posterior is $\log(\zeta \int_X P(Y_{\mathcal{O}}, X, \theta; D, W)dX)$. Because of the conditional independence arising from the design of our likelihood and prior models, we get the complete-data statistical model as

$$P(Y_{\mathcal{O}}, X, \theta; D, W) := P(Y_{\mathcal{O}}, Y_{\mathcal{M}}, X, R, A, \Phi, F; D, W)$$

$$\propto P(Y_{\mathcal{O}}, Y_{\mathcal{M}}, X, A, F|R, \Phi; D, W)$$
$$= P(Y_{\mathcal{O}}|X)P(Y_{\mathcal{M}}|X)P(X|R, A, \Phi, F; D, W)P(A)P(F). \quad (6)$$

The k-t undersampling is equivalent to projecting the full k-space data from a higher-dimensional space to a lower-dimensional subspace. A natural model for the missing k-space values $Y_{\mathcal{M}}$ is i.i.d. Gaussian, like the observed k-space data $Y_{\mathcal{O}}$, having the same acquisition noise variance σ^2. This model is also motivated by the principle of maximum entropy and least prior information [43] when the variance σ^2 of acquisition noise is specified. Thus,

$$P(Y_{\mathcal{M}}|X) \propto \exp\Bigg(-\sum_{k=1}^{K}\sum_{n=1}^{N}\sum_{t=1}^{T}(0.5/\sigma^2) \|Y_{\mathcal{M},n}^{(k,t)} - \mathcal{S}_{\mathcal{M},n}^{t}\mathcal{F}\sum_{l=1+(n-1)M}^{nM}\mathcal{C}_l\Psi_l^k X_l^t\|_2^2\Bigg). \quad (7)$$

E Step: Nested Minorization of Incomplete-Data Log-Posterior. Let θ^i be the parameter estimate at iteration i. The E step minorizes the incomplete-data log-posterior using the minorizer $\mathcal{Q}(\theta;\theta^i) := \mathbb{E}_{P(X|Y_{\mathcal{O}},\theta^i;D,W)}[\log P(Y_{\mathcal{O}}, X, \theta; D, W)]$. The M step later maximizes $\mathcal{Q}(\theta;\theta^i)$ giving $\theta^{(i+1)}$. The non-concavity of $\mathcal{Q}(\theta;\theta^i)$ makes its maximization intractable. We propose nested minorization [40] to further minorize $\mathcal{Q}(\theta;\theta^i)$ by minorizing its expectand $\log P(Y_{\mathcal{O}}, X, \theta; D, W) = \log P(Y_{\mathcal{O}}, Y_{\mathcal{M}}, X|R, A, \Phi, F; D, W) + \log(P(A)P(F))$ using tractable concave minorizers for these two terms as derived in Appendix A and Appendix B, respectively as

$$M(Y_{\mathcal{O}}, Y_{\mathcal{M}}, X|R, A, \Phi, F; D, W) :=$$
$$-\sum_{k=1}^{K}\sum_{n=1}^{N}\sum_{t=1}^{T}\frac{\|Y_n^{(k,t)} - \mathcal{F}\sum_{l=1+(n-1)M}^{nM}\mathcal{C}_l\Psi_l^k X_l^t\|_2^2}{2\sigma^2}$$
$$-\sum_{l=1}^{L}\sum_{v=1}^{V}\delta b_{(l,v)}\|X_{(l,v)} - \Phi_{(l,v)}(RDa_{(l,v)} + d_0 a_{0(l,v)})\|_2^2$$
$$-\sum_{l=1}^{L}\eta\|X_l - Wf_l\|_2^2 + \text{log-normalizing-constant}; \quad (8)$$

$$M(A, F) := -\sum_{l=1}^{L}\sum_{v=1}^{V}\sum_{j=1}^{J}\alpha c_{(l,v)}^{j}(a_{(l,v)}^{j})^2$$
$$-\sum_{l=1}^{L}\sum_{v=1}^{V}\sum_{j=1}^{J}\sum_{(\boldsymbol{l}',\boldsymbol{v}')\in\mathcal{N}_{(l,v)}^{A}}\beta h_{(l,v),(\boldsymbol{l}',\boldsymbol{v}')}^{j}(a_{(l,v)}^{j} - a_{(\boldsymbol{l}',\boldsymbol{v}')}^{j})^2$$
$$-\sum_{l=1}^{L}\rho\|f_l\|_{1,\epsilon}^{1} + \text{log-normalizing-constant}, \quad (9)$$

with nested-minorization-related constants

$$b_{(l,v)} := \frac{p}{2}(\|\widehat{X}_{(l,v)}^{i} - \Phi_{(l,v)}^{i}(R^i Da_{(l,v)}^{i} + d_0 a_{0(l,v)}^{i})\|_2^2 + \epsilon)^{\frac{p}{2}-1}; \quad (10)$$

$$c_{(l,v)}^{j} := \frac{q}{2}((a_{(l,v)}^{j^i})^2 + \epsilon)^{\frac{q}{2}-1}; \quad (11)$$

$$h_{(l,v),(\boldsymbol{l}',\boldsymbol{v}')}^{j} = 0.5\omega_{(l,v),(\boldsymbol{l}',\boldsymbol{v}')} \text{ when } |a_{(l,v)}^{j^i} - a_{(\boldsymbol{l}',\boldsymbol{v}')}^{j^i}| \leqslant \gamma, \quad (12)$$

$$h_{(l,v),(\boldsymbol{l}',\boldsymbol{v}')}^{j} = \frac{\gamma\omega_{(l,v),(\boldsymbol{l}',\boldsymbol{v}')}}{(2((a_{(l,v)}^{j^i} - a_{(\boldsymbol{l}',\boldsymbol{v}')}^{j^i})^2 + \epsilon)^{0.5})} \text{ otherwise.} \quad (13)$$

Thus, the nested minorizer of the incomplete-data log-posterior is $\mathcal{Q}'(\theta;\theta^i) := \mathbb{E}_{P(X|Y_{\mathcal{O}},\theta^i;D,W)}[M(Y_{\mathcal{O}}, Y_{\mathcal{M}}, X|R, A, \Phi, F; D, W) + M(A, F)]$. The first minorizer $\mathcal{Q}(\theta;\theta^i)$ and the nested minorizer $\mathcal{Q}'(\theta;\theta^i)$ both minorize the incomplete-data log posterior. Hence, the minorization-maximization using the nested minorizer also leads to a stationary point.

E Step: Variational Bayesian Modeling. The nested minorization makes the expectand in $\mathcal{Q}(\theta;\theta^i)$ concave, but the expectation as a whole in the nested minorizer $\mathcal{Q}'(\theta;\theta^i)$ remains cumbersome because of the form of the posterior $P(X|Y_{\mathcal{O}}, \theta; D, W)$ of the latent variable X. Thus, we use the variational Bayesian (VB) strategy [44] to first factorize the posterior model $P(X|Y_{\mathcal{O}}, \theta; D, W)$ by $\widehat{P}(X|Y_{\mathcal{O}}, \theta^i; D, W) := \prod_{(l,v),t}\Upsilon_{(l,v)}^{t}(X_{(l,v)}^{t})$, and then propose an (approximate) closed-form representation of the factors by leveraging the concavity of our nested-minorizers of the incomplete-data log-posterior. Our nested minorizer $M(Y_{\mathcal{O}}, Y_{\mathcal{M}}, X|R, A, \Phi, F; D, W)$ can be written as a sum of concave quadratics in X as

$$M(Y_{\mathcal{O}}, Y_{\mathcal{M}}, X|R, A, \Phi, F; D, W) =$$
$$-\sum_{k=1}^{K}\sum_{l=1}^{L}\sum_{v=1}^{V}\sum_{t=1}^{T}\kappa_l^k\|X_{(l,v)}^{t} - m_{(l,v)}^{(k,t)}\|_2^2$$
$$-\sum_{l=1}^{L}\sum_{v=1}^{V}\sum_{t=1}^{T}\delta b_{(l,v)}\|X_{(l,v)}^{t} - g_{(l,v)}^{t}\|_2^2$$
$$-\sum_{l=1}^{L}\sum_{v=1}^{V}\sum_{t=1}^{T}\eta\|X_{(l,v)}^{t} - w_{(l,v)}^{t}\|_2^2$$
$$+ \text{log-normalizing-constant}, \quad (14)$$

where $\kappa_l^k := 0.5\Psi_l^{k^\top}\Psi_l^k/\sigma^2$, $m_l^{(k,t)} := \Psi_l^{k^{-1}}(\mathcal{F}^* Y_n^{(k,t)} - \sum_{l'\neq l}\mathcal{C}_{l'}\Psi_{l'}^k X_{l'}^t)$, $g_{(l,v)} := \Phi_{(l,v)}(RDa_{(l,v)} + d_0 a_{0(l,v)})$ and $w_l = Wf_l$. Because of the quadratic nature of the nested minorizer $M(Y_{\mathcal{O}}, Y_{\mathcal{M}}, X|R, A, \Phi, F; D, W)$, the individual VB factors become Gaussian to give $\Upsilon_{(l,v)}^{t}(X_{(l,v)}^{t}) \approx$

$$\mathbf{N}\Bigg(X_{(l,v)}^{t}; \frac{\sum_{k=1}^{K}\kappa_l^k m_{(l,v)}^{(k,t)} + \delta b_{(l,v)}g_{(l,v)}^{t} + \eta w_{(l,v)}^{t}}{\sum_{k=1}^{K}\kappa_l^k + \delta b_{(l,v)} + \eta}, \frac{0.5}{\sum_{k=1}^{K}\kappa_l^k + \delta b_{(l,v)} + \eta}\Bigg)$$

$$= \mathbf{N}\left(X^t_{(l,v)}; \mu^t_{(l,v)}, (\sigma^t_{(l,v)})^2\right), \quad (15)$$

where $\mathbf{N}(X; \mu, \sigma^2)$ denotes a Gaussian distribution on X with mean μ and variance σ^2. Thus, the VB model for the nested minorizer of the incomplete-data log-posterior is $\mathcal{Q}''(\theta; \theta^i) = \mathbb{E}_{\widehat{P}(X|Y_{\mathcal{O}}, \theta^i; D, W)}[M(Y_{\mathcal{O}}, Y_{\mathcal{M}}, X|R, A, \Phi, F; D, W) + M(A, F)]$.

M Step: Parameter Optimization. We estimate parameters $\theta^{(i+1)}$, for iteration $i+1$, which maximize the approximate nested minorizer $\mathcal{Q}''(\theta; \theta^i)$ by alternately estimating Φ, $Y_{\mathcal{M}}$, R, A and F iteratively. The optimal phase $\Phi_{(l,v)}$, $\forall(l, v)$, is the $T \times T$ unitary diagonal matrix where the diagonals are the complex phases of $\mu_{(l,v)}$. We compute the optimal $Y_{\mathcal{M}}$ by solving the equation $\nabla_{Y_{\mathcal{M}}} \mathcal{Q}''(\theta; \theta^i) = 0$ as

$$(Y^{(k,t)}_{\mathcal{M},n})^{(i+1)} = \mathcal{S}^t_{\mathcal{M},n} \mathcal{F} \sum_{l=1+(n-1)M}^{nM} \mathcal{C}_l \Psi^k_l (\mu^t_l)^{(i+1)}. \quad (16)$$

We compute the optimal R by maximizing $\mathcal{Q}''(\theta; \theta^n)$ using projected gradient ascent with adaptive learning rate. We compute the optimal A by solving the equation $\nabla_A \mathcal{Q}''(\theta; \theta^i) = 0$. We estimate the optimal F using ADMM-LASSO [45] optimization on $\mathcal{Q}''(\theta; \theta^i)$.

D. SUMMARY OF ALGORITHM FOR R-fMRI IMAGE RECONSTRUCTION

We summarize the algorithm for implementation of our proposed DW+VBEMNM framework for inference of the reconstructed R-fMRI and its uncertainty of reconstruction.

Initialization. We initialize the variables in two stages. In the first stage, we initialize: (i) the missing k-space values $Y^{(k,\tau)}_{\mathcal{M},n}$ for the set τ of acquired timeframes, using GRAPPA [12] on $Y^{(k,\tau)}_{\mathcal{O},n}$; (ii) X^τ for the acquired timeframes by using SENSE-CAIPI [16] on the multicoil data $\mathcal{F}^* Y^{(k,t)}$ $\forall k$, and using linear interpolation on X^τ for the timeframes $t \notin \tau$; (iii) $R \leftarrow I$; (iv) $a^j_{(l,v)}$ as done for dictionary learning; (v) $a_{0(l,v)} = \sum_{t=1}^T |X^t_{(l,v)}|/T$; (vi) $\Phi_{(l,v)}$ as the T × T unitary diagonal matrix where the diagonals are the complex phases of $X_{(l,v)}$. In the second stage, we re-initialize: (i) X by fitting only the dictionary D to it (without wavelets) using gradient ascent with adaptive learning rate as in [7], with $\{R, A\}$ kept constant; (ii) $Y^{(k,t)}_{\mathcal{M},n}$ by using the sampling operator $\mathcal{S}^t_{\mathcal{M},n}$ on the k-space SMS-CAIPI image of corresponding slices in the new estimates of X; (iii) $a^j_{(l,v)}$, $a_{0(l,v)}$ and $\Phi_{(l,v)}$ as in the first stage using the new estimates of X; (iv) $f_l = W^* X_l$.

Parameter Estimation: E Step and M Step. We repeat the following procedure till convergence.

1) *E Step:* We update the mean and variance of the latent variables $X^t_{(l,v)}$ as $(\mu^t_{(l,v)})^{(i+1)} := (\sum_{k=1}^K \kappa^k_l m^{(k,t)}_{(l,v)} + \delta b_{(l,v)} g^t_{(l,v)} + \eta w^t_{(l,v)})/(\sum_{k=1}^K \kappa^k_l + \delta b_{(l,v)} + \eta)$ and $((\sigma^t_{(l,v)})^2)^{(i+1)} := 0.5/(\sum_{k=1}^K \kappa^k_l + \delta b_{(l,v)} + \eta)$ respectively, to get the MAP estimate $(\widehat{X}^t_{(l,v)})^{(i+1)} := (\mu^t_{(l,v)})^{(n+1)}$. Because the value of $((\sigma^t_{(l,v)})^2)$ is independent of t, we denote it simply by $\sigma^2_{(l,v)}$.
2) *M step:* We update parameters $Y_{\mathcal{M}}$, R, A, Φ and F as described before, by alternately optimizing $\mathcal{Q}''$.

Inferring Reconstructed Image and Uncertainty from Latent-Variable Distribution. When the VBEMNM converges to the optimal parameters θ^*, we infer (i) the underlying reconstructed R-fMRI image X as the posterior mode μ^*, and (ii) the per-voxel uncertainty of reconstruction, for each voxel v and time t, through the posterior standard deviation $\sigma_{(l,v)}$.

IV. RESULTS AND DISCUSSION

We evaluate the performance of our reconstruction framework on brain R-fMRI data. We compare our **DW+VBEMNM** framework with other methods of reconstructing R-fMRI from undersampled SMS-CAIPI data when the coil sensitivities and CAIPI shifts are known (or can be reliably estimated) (i) **W+GA**: uses wavelet (W) models akin to those in [9], [46] and optimizes using gradient ascent (GA) with adaptive learning rate; (ii) **GRP+SENSE**: uses GRAPPA (GRP) [12] to estimate the missing SMS k-space data per coil, followed by SENSE+CAIPI [16] to reconstruct individual slices; (iii) **LR+SENSE**: uses a low-rank (LR) model akin to [8] to estimate the full SMS per-coil k-space matrix from the undersampled data (the sampled k-space data is replaced after estimation as in [8] to improve the quality of reconstruction) and SENSE+CAIPI [16] to reconstruct individual slices.

Because GRAPPA and the low-rank modeling methods do *not* have explicit mechanisms for temporal regularization, unlike the temporal dictionary in our DW+VBEMNM or the spatiotemporal wavelet in our DW+VBEMNM and W+GA, GRP+SENSE and LR+SENSE cannot be used to reconstruct from temporally-undersampled data. Thus, we extend them to (i) **GRP+SENSE-T**: uses GRP+SENSE to first fill-in the missing k-space data, then reconstruct individual slices for all the timeframes for which some data was acquired, and finally uses cubic interpolation to reconstruct the missing timeframes; (ii) **LR+SENSE-T**: uses LR+SENSE to first fill-in the missing k-space data, then reconstruct individual slices for all the timeframes for which some data was acquired, and finally uses cubic interpolation to reconstruct the missing timeframes. We measure the performance in terms of (i) the mean structural similarity (mSSIM) [47] between the ground truth and the reconstructed RSNs, and (ii) GM-tSNR of the reconstructed image, where we extend the tSNR measure in [48] (typically used for R-fMRI; measures the average over all gray-matter voxels of the ratio of mean of time-series to the standard deviation of the time-series) to be applicable to complex-valued images, by replacing the mean of the time-series with its root-mean-square. We estimate the RSNs using standard seed-based normalized cross-correlations.

A. RESULTS ON BRAIN R-fMRI DATA

We compare the performance of all methods on high-quality brain R-fMRI data from the Human Connectome Project

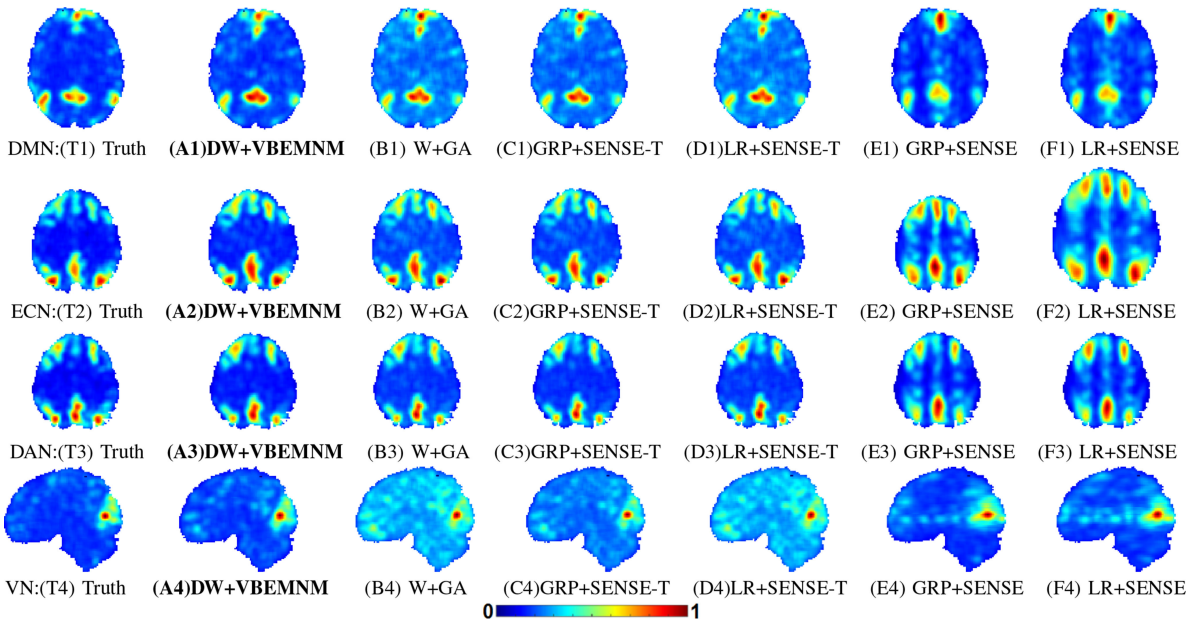

FIGURE 1. Results on Brain R-fMRI Data. ***DMN*** **estimated from (T1) original data; from reconstructions of k-t undersampled SMS-CAIPI data obtained by (A1) our DW+VBEMNM: mSSIM 0.97; (B1) W+GA: mSSIM 0.92; (C1) GRP+SENSE-T: mSSIM 0.93; (D1) LR+SENSE-T: mSSIM 0.92; (E1) GRP+SENSE: mSSIM 0.94; (F1) LR+SENSE: mSSIM 0.90.** ***ECN*** **estimated from (T2) original data; from reconstructions of k-t undersampled SMS-CAIPI data obtained by (A2) our DW+VBEMNM: mSSIM 0.96; (B2) W+GA: mSSIM 0.92; (C2) GRP+SENSE-T: mSSIM 0.93; (D2) LR+SENSE-T: mSSIM 0.92; (E2) GRP+SENSE: mSSIM 0.85; (F2) LR+SENSE: mSSIM 0.84.** ***DAN*** **estimated from (T3) original data; from reconstructions of k-t undersampled SMS-CAIPI data obtained by (A3) our DW+VBEMNM: mSSIM 0.97; (B3) W+GA: mSSIM 0.94; (C3) GRP+SENSE-T: mSSIM 0.95; (D3) LR+SENSE-T: mSSIM 0.94; (E3) GRP+SENSE: mSSIM 0.90; (F3) LR+SENSE: mSSIM 0.90.** ***VN*** **estimated from (T4) original data; from reconstructions of k-t undersampled SMS-CAIPI data obtained by (A4) our DW+VBEMNM: mSSIM 0.96; (B4) W+GA: mSSIM 0.78; (C4) GRP+SENSE-T: mSSIM 0.89; (D4) LR+SENSE-T: mSSIM 0.78; (E4) GRP+SENSE: mSSIM 0.94; (F4) LR+SENSE: mSSIM 0.93.**

(HCP). The HCP data has $2\times2\times2$ mm^3 voxels, approximately 1.4 Hz temporal sampling rate and 1200 timeframes (approximately 15 minutes scan), which is much higher than the ideal Nyquist rate of 0.2 Hz for R-fMRI time-series. We use the dictionary learned from the fully-sampled, non-SMS ground truth data of one subject to reconstruct R-fMRI of 50 other evaluation subjects from their corrupted k-t undersampled SMS-CAIPI data. The ground truth data has a GM-tSNR of about 76.1, averaged across all evaluation subjects.

We use coil sensitivity profiles obtained from a Siemens mMR scanner to generate multicoil SMS data for 32 coils with $5\times$ SMS and CAIPI. We retrospectively undersample the data $1.5\times$ in k-space and $4\times$ in time to give $6\times$ k-t undersampling for our DW+VBEMNM, W+GA, GRP+SENSE-T, and LR+SENSE-T. Because GRP+SENSE and LR+SENSE do *not* have explicit mechanisms for temporal regularization, we retrospectively undersample the data purely in k-space. However, we maintain the total undersampling factor by undersampling $6\times$ in k-space for GRP+SENSE and LR+SENSE for a fair comparison in terms of the speedup factor. Our k-space undersampling pattern can be easily implemented on commercial scanners by sampling line encodes densely near the k-space center and equidistantly elsewhere. Our framework can be extended to 3D k-space undersampling by modifying the likelihood term with the same priors. We introduce i.i.d. additive complex-Gaussian noise, with a small standard deviation, to the undersampled k-space data.

For the ACS for GRP+SENSE-T and GRP+SENSE, we use the fully-sampled first 50 SMS-CAIPI timeframes of the evaluation subject being reconstructed. We set $p = 0.5$ and $q = 0.9$ for our DW+VBEMNM, and tune free parameters for all methods using cross-validation on a validation set (separate from the training subject and evaluation dataset) to give the best reconstructions on the entire evaluation set. We evaluate all methods based on the mSSIM between the ground-truth RSNs (obtained from the fully-sampled R-fMRI data) and the reconstructed RSNs for the default mode network (DMN), executive control network (ECN) and the dorsal attentive network (DAN) and visual network (VN).

RSNs reconstructed using our DW+VBEMNM (Fig. 1(A1)–(A4)) are qualitatively and quantitatively much closer to the ground truth (Fig. 1(T1)–(T4)) RSNs in terms of the structure, lumination, and contrast that are the key components underlying the mSSIM measure (Table 1). They show high structural similarity and negligible artifacts. Reconstructions from W+GA (Fig. 1(B1)–(B4)), GRP+SENSE-T (Fig. 1(C1)–(C4)), and LR+SENSE-T (Fig. 1(D1)–(D4)) are prone to artifacts in the form of spurious high correlations in regions outside those indicate by the ground-truth RSNs. Reconstructions from GRP+SENSE (Fig. 1(E1)–(E4)) and LR+SENSE (Fig. 1(F1)–(F4)) lead to RSNs that show significant deviation from the ground truth RSNs, as expected. Hence, we exclude them from further analysis.

TABLE 1. Results on Brain R-fMRI Data. Average mSSIM (and Standard Deviation) of RSNs for k-t Undersampled SMS-CAIPI Data Over 50 Evaluation Subjects.

Method	Resting-State Networks				Average
	DMN	ECN	DAN	VN	
DW+VBEMNM	**0.97 (0.02)**	**0.96 (0.02)**	**0.95 (0.03)**	**0.96 (0.01)**	**0.96 (0.02)**
W+GA	0.94 (0.02)	0.93 (0.04)	0.92 (0.06)	0.85 (0.05)	0.91 (0.06)
GRP+SENSE-T	0.96 (0.02)	0.95 (0.03)	0.93 (0.05)	0.94 (0.03)	0.94 (0.04)
LR+SENSE-T	0.94 (0.02)	0.93 (0.04)	0.92 (0.06)	0.85 (0.05)	0.91 (0.06)
GRP+SENSE	0.93 (0.01)	0.92 (0.04)	0.89 (0.06)	0.93 (0.02)	0.92 (0.04)
LR+SENSE	0.93 (0.02)	0.91 (0.04)	0.87 (0.07)	0.89 (0.04)	0.90 (0.05)

TABLE 2. Results on Brain R-fMRI Data. Average GM-tSNR and Standard Deviation for Reconstructions from k-t Undersampled SMS-CAIPI Data Over 50 Evaluation Subjects.

Method	Average GM-tSNR	Standard Deviation
Ground Truth	76.1	6.5
DW+VBEMNM	**172.7**	**8.4**
W+GA	86.7	3.6
GRP+SENSE-T	88.9	3.5
LR+SENSE-T	86.7	3.6

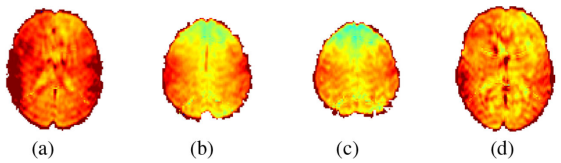

FIGURE 2. Uncertainty Maps Corresponding to Orientation of Acquired Slices. Uncertainty in reconstruction within: (a) DMN region, (b) ECN region, (c) DAN region, and (d) VN region.

Comparing the results from GRP+SENSE and LR+SENSE to those from GRP+SENSE-T and LR+SENSE-T respectively, we observe that although these methods primarily focus on k-space filling, their performance is better when the k-space undersampling factor is lower than the temporal undersampling factor, while maintaining the total (product) k-t undersampling factor. Because our DW+VBEMNM reconstruction framework employs a rich temporal prior model (dictionary) on the BOLD time-series, it is able to leverage larger temporal undersampling factors, instead of larger k-space undersampling factors. Similarly, because W+GA employs a spatiotemporal wavelet model as a prior, and does *not* learn an explicit mechanism for k-space filling, it also works best when the k-space undersampling factor is low. Thus, for a specified speedup factor, all methods considered for further analysis, i.e., DW+VBEMNM, W+GA, GRP+SENSE-T, and LR+SENSE-T, perform best when the k-space undersampling factors are lower allowing higher temporal undersampling factors. Thus, we propose undersampling in k-space by $1.5\times$ and in time by $4\times$.

B. TEMPORAL STABILITY OF RECONSTRUCTED IMAGES

Apart from giving better RSNs than other methods, DW+VBEMNM reconstructions are also significantly superior in terms of the average GM-tSNR that is a standard measure of temporal stability [49]. All methods for reconstruction from k-t undersampled SMS-CAIPI data lead to reconstructed images having higher average GM-tSNR than the ground truth (Table 2). At the existing voxel sizes in the ground truth, our DW+VBEMNM more than doubles the average GM-tSNR over the ground truth to 172.7. Because GM-tSNR reduces with smaller voxel sizes [50], this ability of our DW+VBEMNM makes it a suitable framework for reconstruction with the aim of imaging at higher spatial resolution. Reconstructions from other methods have very low average GM-tSNR, compared to our DW+VBEMNM (Table 2).

C. UNCERTAINTY ESTIMATION

Unlike other methods, in addition to the reconstructed 4D spatiotemporal image, our DW+VBEMNM is able to provide per-voxel uncertainty associated with the reconstructed image. We define the uncertainty of reconstruction at a given voxel to be the ratio of the standard deviation $\sigma_{(l,v)}$ of reconstructed time-series as obtained in Section III to the norm of the reconstructed R-fMRI time-series at voxel (l, v). This definition of uncertainty helps to interpret the standard deviation at a given voxel in the context of its time-series. We find that the spatial pattern of uncertainty indicates higher values at regions consistent with the pattern of undersampling artifacts associated with the k-space undersampling in the transaxial plane and at the pial surface of the cortex (Fig. 2).

D. ENABLING HIGHER SPATIAL RESOLUTION

Our framework proposes a $30\times$ speedup through a $5\times$ speedup stemming from SMS imaging coupled with $6\times$ speedup from k-t undersampling (i.e., $1.5\times$ k-space undersampling and $4\times$ time undersampling). This speedup can potentially allow a $3\times$ reduction in voxel width along each of the three spatial dimensions, specially since the $4\times$ lower temporal sampling rate is still above the ideal Nyquist rate of 0.2 Hz, without increasing the total scan time and, consequently, discomfort for the patient. Further, the ability of our framework to significantly boost the GM-tSNR can help to counter the effects of higher noise at higher spatial resolutions.

We demonstrate the benefits of the potential spatial resolution improvement using our framework through an equivalent experiment. Consider a fully-sampled ground-truth R-fMRI acquisition with (large) $6\times6\times6$ mm^3 voxels and 1.4 Hz temporal sampling rate that utilizes full 15 minutes of total scan time. Because of the larger voxel size, this low-spatial-resolution acquisition has a lower noise level with an average GM-tSNR of 125.6 with a standard deviation of 11.6 over

TABLE 3. Ablation Study. Average mSSIM (and Standard Deviation) of RSNs for k-t Undersampled SMS-CAIPI Data Over 50 Evaluation Subjects.

Method	Resting-State Networks				Average
	DMN	ECN	DAN	VN	
DW+VBEMNM	**0.97 (0.02)**	**0.96 (0.01)**	**0.95 (0.03)**	**0.96 (0.01)**	**0.96 (0.02)**
D+VBEMNM	0.96 (0.01)	0.95 (0.02)	0.94 (0.05)	0.88 (0.03)	0.93 (0.04)
D+PGA	0.95 (0.02)	0.95 (0.04)	0.93 (0.05)	0.94 (0.03)	0.94 (0.04)
BDW+VBEMNM	**0.97 (0.02)**	**0.96 (0.01)**	**0.95 (0.03)**	**0.96 (0.01)**	**0.96 (0.02)**

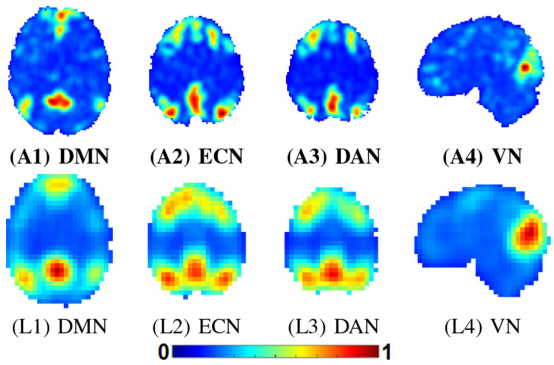

FIGURE 3. Enabling Higher Spatial Resolution. DMN estimated from (A1) reconstructions of k-t undersampled SMS-CAIPI data using our DW+VBEMNM; (L1) 27× low spatial resolution original data. ECN estimated from (A2) reconstructions of k-t undersampled SMS-CAIPI data using our DW+VBEMNM; (L2) 27× low spatial resolution original data. DAN estimated from (A3) reconstructions of k-t undersampled SMS-CAIPI data using our DW+VBEMNM; (L3) 27× low spatial resolution original data. VN estimated from (A4) reconstructions of k-t undersampled SMS-CAIPI data using our DW+VBEMNM; (L4) 27× low spatial resolution original data. Ground Truth RSNs same as in Fig. 1.

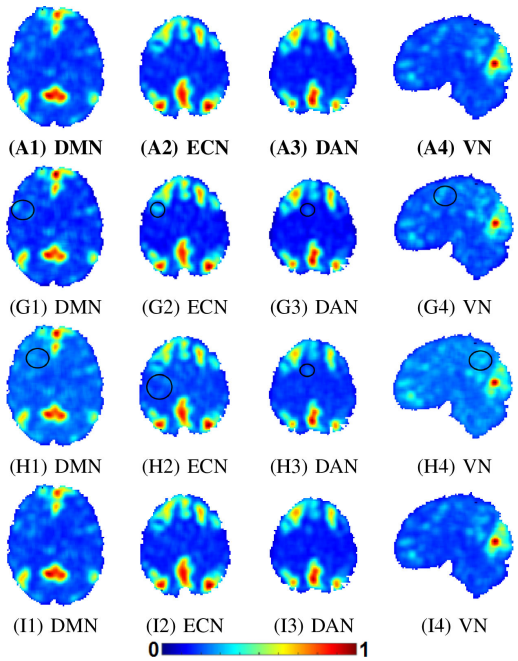

FIGURE 4. Ablation Study. DMN estimated from reconstructions of k-t undersampled SMS-CAIPI data using (A1) DW+VBEMNM: mSSIM 0.97; (G1) D+VBEMNM: mSSIM 0.95; (H1) D+PGA: mSSIM 0.93; (I1) BDW+VBEMNM: mSSIM 0.97. ECN estimated from reconstructions of k-t undersampled SMS-CAIPI data using (A2) DW+VBEMNM: mSSIM 0.96; (G2) D+VBEMNM: mSSIM 0.95; (H2) D+PGA: mSSIM 0.93; (I2) BDW+VBEMNM: mSSIM 0.96. DAN estimated from reconstructions of k-t undersampled SMS-CAIPI data using (A3) DW+VBEMNM: mSSIM 0.97; (G3) D+VBEMNM: mSSIM 0.95; (H3) D+PGA: mSSIM 0.95; (I3) BDW+VBEMNM: mSSIM 0.97. VN estimated from reconstructions of k-t undersampled SMS-CAIPI data using (A4) DW+VBEMNM: mSSIM 0.96; (G4) D+VBEMNM: mSSIM 0.85; (H4) D+PGA: mSSIM 0.88; (I4) BDW+VBEMNM: mSSIM 0.96. Ground Truth RSNs same as in Fig. 1.

all evaluation subjects. We reduce the number of acquired measurements through 5× SMS imaging, 1.5× kspace undersampling, and 4× time undersampling. If we keep the total scan duration unchanged (i.e., 15 minutes), then we can utilize the freed-up time (because of the reduced number of measurements) to increase spatial resolution as follows. We propose to increase the spatial resolution by 27× (3× along each spatial dimension) to $2\times2\times2$ mm^3 and reconstruct the fully-sampled high spatiotemporal resolution signal. Such an acquisition can be practically feasible considering all time constraints. For reconstruction frameworks other than our DW+VBEMNM, this high spatial resolution reconstruction would have (i) lower GM-tSNR, because of the small voxel size [50], as indicated by the GM-tSNR values of reconstructions from these other methods in Section IV-B, and (ii) poorer RSNs as shown in Section IV-A. Nevertheless, despite the higher spatial resolution, our DW+VBEMNM can still maintain a high GM-tSNR, as seen in Section IV-B, beyond the GM-tSNR of the lower spatial resolution ground truth. This ability of our DW+VBEMNM, coupled with the ability to improve RSN estimation, makes it a better framework for enabling higher spatial resolution. We compare the RSNs estimated from the low spatial resolution acquisition and our high spatial resolution reconstruction. RSNs estimated from DW+VBEMNM are clearly better than those from the low-spatial-resolution acquisition (Fig. 3). Analogous to this experiment, we can potentially improve the spatial resolution from the current voxel size of $2\times2\times2$ mm^3 voxels to a voxel size of $0.67\times0.67\times0.67$ mm^3.

E. ABLATION STUDY

To justify the contributions of the novel features of joint dictionary+wavelet modeling and the VBEMNM optimization strategy (Fig. 4(A1)–(A4), Table 3, Table 4) to our reconstruction framework, we compare the performance of our proposed DW+VBEMNM framework with its ablated prior variants. We find that our dictionary-based methods outperform other methods. Within our dictionary-based methods, our

TABLE 4. Ablation Study. Average GM-tSNR and Standard Deviation for Reconstructions From k-t Undersampled SMS-CAIPI Data Over 50 Evaluation Subjects.

Method	Average GM-tSNR	Standard Deviation
DW+VBEMNM	**172.7**	**8.4**
D+VBEMNM	84.5	4.8
D+PGA	93.1	3.4
BDW+VBEMNM	**172.7**	**8.4**

DW+VBEMNM performs better than others. Reconstructed RSNs using only our dictionary model (without wavelets) with VBEMNM optimization (i.e., D+VBEMNM), which is similar to our work in [40], leads to qualitatively and quantitatively poorer RSNs (Fig. 4(G1)–(G4), Table 3). Specially, unlike our DW+VBEMNM, the D+VBEMNM reconstruction is unable to boost the GM-tSNR significantly (Table 4). This might reduce its utility for enabling higher spatial resolution because the temporal stability measured in terms of the GM-tSNR reduces significantly at higher spatial resolutions as demonstrated in Section IV-D. Further, replacing the VBEMNM optimization strategy with projected-gradient-ascent (PGA) optimization for our dictionary model (i.e., D+PGA), as presented in our work in [7], leads to some spurious high correlations in the RSNs (Fig. 4(H1)–(H4),Table 3) and a lower GM-tSNR (Table 4). We further justify the adaptability of our dictionary learned from one training subject to other evaluation subjects using similarity transforms through blind DW+VBEMNM (BDW+VBEMNM). We reconstruct RSNs using our dictionary+wavelet model and VBEMNM optimization where the dictionary D is also updated in the VBEMNM optimization. We find that updating the dictionary does *not* lead to any observable difference in the RSNs and reconstruction quality (Fig. 4(I1)–(I4),Table 3, Table 4) which indicates that our similarity transform is enough to adapt the dictionary to the evaluation subject, and approaches like blind compressive sensing [28] are *not* required.

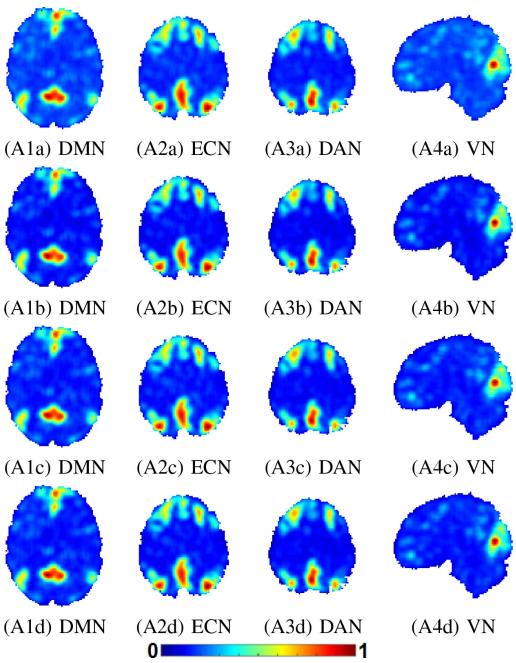

FIGURE 5. Evaluating Sensitivity to Training Data. *DMN* estimated from reconstructions of k-t undersampled SMS- CAIPI data via our DW+VBEMNM with (A1a) Dictionary 1: mSSIM 0.967, (A1b) Dictionary 2: mSSIM 0.972, (A1c) Dictionary 3: mSSIM 0.966, (A1d) Dictionary 4: mSSIM 0.969. *ECN* estimated from reconstructions of k-t undersampled SMS- CAIPI data via our DW+VBEMNM withh (A2a) Dictionary 1: mSSIM 0.961, (A2b) Dictionary 2: mSSIM 0.966, (A2c) Dictionary 3: mSSIM 0.952, (A2d) Dictionary 4: mSSIM 0.964. *DAN* estimated from reconstructions of k-t undersampled SMS- CAIPI data via our DW+VBEMNM with (A3a) Dictionary 1: mSSIM 0.970, (A3b) Dictionary 2: mSSIM 0.971, (A3c) Dictionary 3: mSSIM 0.962, (A3d) Dictionary 4: mSSIM 0.971. *VN* estimated from reconstructions of k-t undersampled SMS- CAIPI data via our DW+VBEMNM with (A4a) Dictionary 1: mSSIM 0.959, (A4b) Dictionary 2: mSSIM 0.978, (A4c) Dictionary 3: mSSIM 0.961, (A4d) Dictionary 4: mSSIM 0.964.

F. EVALUATING SENSITIVITY TO SELECTION OF TRAINING DATA

We learn three more dictionaries from the fully-sampled non-SMS ground truth data of three different subjects that are distinct from the first training subject and the evaluation subjects. We compare RSNs estimated from reconstructions obtained using each of these dictionaries in our DW+VBEMNM framework. We find negligible qualitative difference (Fig. 5) in the RSNs estimated from reconstructions obtained using different dictionaries. The change in average mSSIM over all evaluation subjects is also insignificant (Table 5).

G. EVALUATING SENSITIVITY TO HEAD MOTION AND TIME LAG

We evaluate the performance of all methods in the presence of practical complications like artifacts arising due to (i) head motion and (ii) time lag in the acquisition of slices arising due to reduced temporal sampling rate.

We mimic head motion in the evaluation data based on the process detailed in [51], where the head rotates about the spine (transverse axis). The head oscillates from θ_{angle} to the left to θ_{angle} to the right every two minutes in the 15-minute scan. The angle θ_{angle} is selected to mimic realistic head motion as suggested in [51]. We generate SMS-CAIPI data from this motion-induced data, undersample in k-space, and introduce noise as described in Section IV-A.

We propose to increase the spatial resolution in R-fMRI acquisition through SMS-CAIPI imaging and k-t undersampling. Specifically, in this paper, we propose $4\times$ undersampling in time by reducing the temporal sampling rate by $4\times$. This means that one timeframe of the R-fMRI should be acquired over $4\times$ the current time interval between consecutive timeframes. This would lead to an increase in the time lag as acquisition proceeds from slices in the superior to the slices in the inferior region of the brain. We incorporate this time lag into the data as follows. Assuming that, in the ground truth, each entire (thicker) slice is acquired at a specific timepoint,

TABLE 5. Evaluating Sensitivity to Selection of Training Data. Average mSSIM (and Standard Deviation) of RSNs for k-t Undersampled SMS-CAIPI Data Reconstructed Using Dictionaries Learned From Different Subjects and our DW+VBEMNM Framework.

Method	Resting-State Networks				Average
	DMN	ECN	DAN	VN	
Dictionary 1	0.968 (0.016)	0.961 (0.017)	0.953 (0.031)	0.962 (0.014)	0.961 (0.020)
Dictionary 2	0.975 (0.010)	0.964 (0.017)	0.965 (0.022)	0.968 (0.013)	0.968 (0.016)
Dictionary 3	0.968 (0.015)	0.958 (0.021)	0.950 (0.033)	0.964 (0.013)	0.960 (0.022)
Dictionary 4	0.969 (0.016)	0.962 (0.018)	0.956 (0.026)	0.967 (0.011)	0.964 (0.019)

TABLE 6. Evaluating Sensitivity to Head Motion and Time Lag. Average mSSIM (and Standard Deviation) of RSNs for k-t Undersampled SMS-CAIPI Data With Head Motion and Time Lag Over 50 Evaluation Subjects.

Method	Resting-State Networks				Average
	DMN	ECN	DAN	VN	
DW+VBEMNM	**0.95 (0.02)**	**0.93 (0.04)**	**0.92 (0.06)**	**0.92 (0.02)**	**0.93 (0.04)**
W+GA	0.90 (0.06)	0.87 (0.06)	0.88 (0.08)	0.89 (0.03)	0.89 (0.06)
GRP+SENSE-T	0.92 (0.04)	0.90 (0.05)	0.90 (0.07)	0.91 (0.02)	0.91 (0.05)
LR+SENSE-T	0.90 (0.06)	0.87 (0.06)	0.88 (0.08)	0.89 (0.03)	0.89 (0.06)

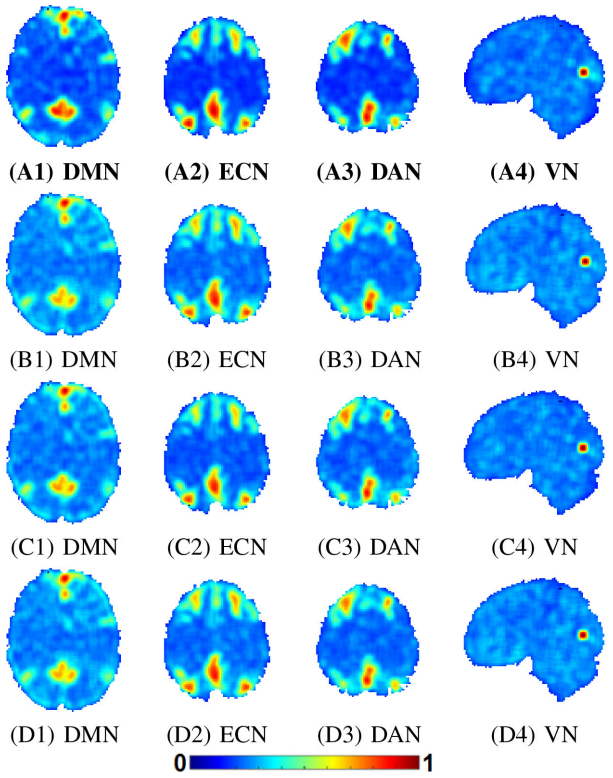

FIGURE 6. Evaluating Sensitivity to Head Motion and Time Lag. DMN estimated from reconstructions of k-t undersampled SMS-CAIPI data with head motion and time lag artifacts, obtained by (A1) DW+VBEMNM: mSSIM 0.94; (B1) W+GA: mSSIM 0.89; (C1) GRP+SENSE-T: mSSIM 0.91; (D1) LR+SENSE-T: mSSIM 0.89. ECN estimated from reconstructions of k-t undersampled SMS-CAIPI data with head motion and time lag artifacts, obtained by (A2) DW+VBEMNM: mSSIM 0.94; (B2) W+GA: mSSIM 0.87; (C2) GRP+SENSE-T: mSSIM 0.91; (D2) LR+SENSE-T: mSSIM 0.87. DAN estimated from reconstructions of k-t undersampled SMS-CAIPI data with head motion and time lag artifacts, obtained by (A3) DW+VBEMNM: mSSIM 0.93; (B3) W+GA: mSSIM 0.87; (C3) GRP+SENSE-T: mSSIM 0.87; (D3) LR+SENSE-T: mSSIM 0.87. VN estimated from reconstructions of k-t undersampled SMS-CAIPI data with head motion and time lag artifacts, obtained by (A4) DW+VBEMNM: mSSIM 0.89; (B4) W+GA: mSSIM 0.85; (C4) GRP+SENSE-T: mSSIM 0.87; (D4) LR+SENSE-T: mSSIM 0.85. Ground Truth RSNs same as in Figure 1.

we simulate the intensities in each (thinner) slice in the higher-spatial-resolution scan (that has a larger number of slices) using linear interpolation between the thicker slices, based on the timepoint at which the thinner slice would be acquired.

We find that our DW+VBEMNM continues to outperform other methods for reconstruction from k-t undersampled SMS-CAIPI R-fMRI data qualitatively and quantitatively even in the presence of head motion and time lag (Fig. 6, Table 6).

V. CONCLUSION

We demonstrated a novel algorithmic framework to reconstruct R-MRI from k-t undersampled SMS-CAIPI data, unlike other methods that use only k-space undersampling and SMS. We propose a coupled prior model, incorporating (i) a robust spatially-regularized temporal-dictionary prior and (ii) a spatiotemporal wavelet prior, which we fit efficiently using variational Bayesian expectation maximization with nested minorization. We describe our optimization algorithm to handle the non-concavity of our robust model using nested minorization within VBEM, giving efficient parameter updates within each iteration of EM. Our framework enables uncertainty estimation, and we find that the spatial pattern of uncertainty indicates higher values at regions consistent with the pattern of undersampling artifacts associated with the k-space undersampling and at the pial surface of the cortex. RSN estimates, from 5× SMS 6× k-t undersampled data (total 30× speedup), from our framework, compare favourably to those from existing methods for R-fMRI reconstruction. Further, our framework significantly boosts the GM-tSNR of the reconstruction. We demonstrated the insensitivity of our framework to dictionary models learned from different subjects. We also demonstrated that our framework is less sensitive to head motion and time-lag artifacts as compared to other methods. In these ways, our framework can potentially enable 3× isotropic improvement in spatial resolution (total 27× higher spatial resolution) in R-fMRI, even after accounting for practical setup and acquisition times in R-fMRI acquisition,

without increasing the current typical scan time. The current framework may be limited in its ability to handle severe motion artifacts. To address this issue, the current framework can be extended with sophisticated head-motion modeling in future work.

APPENDIX A
DERIVATION FOR NESTED MINORIZER $M(Y_{\mathcal{O}}, Y_{\mathcal{M}}, X|R, A, \Phi, F; D, W)$

We have

$$\begin{aligned}
&\log P(Y_{\mathcal{O}}, Y_{\mathcal{M}}, X|R, A, \Phi, F; D, W) := \\
&-\sum_{k=1}^{K}\sum_{n=1}^{N}\sum_{t=1}^{T}\frac{\|Y_n^{(k,t)} - \mathcal{F}\sum_{l=1+(n-1)M}^{nM}\mathcal{C}_l\Psi_l^k X_l^t\|_2^2}{2\sigma^2} \\
&-\sum_{l=1}^{L}\sum_{v=1}^{V}\delta\|X_{(l,v)} - \Phi_{(l,v)}(RDa_{(l,v)} + d_0 a_{0(l,v)})\|_{2,\epsilon}^p \\
&-\sum_{l=1}^{L}\eta\|X_l - Wf_l\|_2^2 + \text{log-normalizing-constant.} \qquad (17)
\end{aligned}$$

We choose the previous i-th iteration $(\widehat{X}^i, R^i, A^i, \Phi^i)$ estimate as the point where the minorizer $M(Y_O, Y_M, X|R, A, \Phi, F; D, W)$ touches the original function $\log P(Y_O, Y_M, X|R, A, \Phi, F; D, W)$. We design the minorizer to be of concave quadratic nature as

$$\begin{aligned}
&M(Y_{\mathcal{O}}, Y_{\mathcal{M}}, X|R, A, \Phi, F; D, W) := \\
&-\sum_{k=1}^{K}\sum_{n=1}^{N}\sum_{t=1}^{T}\frac{\|Y_n^{(k,t)} - \mathcal{F}\sum_{l=1+(n-1)M}^{nM}\mathcal{C}_l\Psi_l^k X_l^t\|_2^2}{2\sigma^2} \\
&-\sum_{l=1}^{L}\sum_{v=1}^{V}\delta b_{(l,v)}\|X_{(l,v)} - \Phi_{(l,v)}(RDa_{(l,v)} + d_0 a_{0(l,v)})\|_2^2 \\
&-\sum_{l=1}^{L}\eta\|X_l - Wf_l\|_2^2 + \text{log-normalizing-constant.} \qquad (18)
\end{aligned}$$

We equate the values and gradients of $\log P(Y_O, Y_M, X|R, A, \Phi, F; D, W)$ and $M(Y_O, Y_M, X|R, A, \Phi, F; D, W)$ at the touching point $(\widehat{X}^i, R^i, A^i, \Phi^i)$ to get

$$b_{(l,v)} := \frac{p}{2}(\|\widehat{X}^i_{(l,v)} - \Phi^i_{(l,v)}(R^i D a^i_{(l,v)} + d_0 a^i_{0(l,v)})\|_2^2 + \epsilon)^{\frac{p}{2}-1}. \qquad (19)$$

and the constant $b'_{(l,v)}$. However, $b'_{(l,v)}$ is absorbed into the log-normalizing-constant, because it is an additive constant, and therefore need *not* be optimized.

APPENDIX B
DERIVATION FOR NESTED MINORIZER $M(A, F)$

We have

$$\begin{aligned}
&\log P(A, F) := -\sum_{l=1}^{L}\sum_{v=1}^{V}\alpha\|a_{(l,v)}\|_{q,\epsilon}^q \\
&-\sum_{l=1}^{L}\sum_{v=1}^{V}\sum_{j=1}^{J}\sum_{(\boldsymbol{l}',\boldsymbol{v}')\in\mathcal{N}^A_{(l,v)}}\beta\omega_{(l,v),(\boldsymbol{l}',\boldsymbol{v}')}H_\gamma(a^j_{(l,v)} - a^j_{(\boldsymbol{l}',\boldsymbol{v}')}) \\
&-\sum_{l=1}^{L}\rho\|f_l\|_{1,\epsilon}^1 + \text{log-normalizing-constant.} \qquad (20)
\end{aligned}$$

We choose the previous i-th iteration (A^i, F^i) estimate as the point where the minorizer $M(A, F)$ touches the original function $\log P(A, F)$. We design the minorizer to be of concave nature as

$$\begin{aligned}
&M(A, F) := -\sum_{l=1}^{L}\sum_{v=1}^{V}\sum_{j=1}^{J}c^j_{(l,v)}(a^j_{(l,v)})^2 \\
&-\sum_{l=1}^{L}\sum_{v=1}^{V}\sum_{j=1}^{J}\sum_{(\boldsymbol{l}',\boldsymbol{v}')\in\mathcal{N}^A_{(l,v)}}\beta h^j_{(l,v),(\boldsymbol{l}',\boldsymbol{v}')}(a^j_{(l,v)} - a^j_{(\boldsymbol{l}',\boldsymbol{v}')})^2 \\
&-\sum_{l=1}^{L}\rho\|f_l\|_{1,\epsilon}^1 + \text{log-normalizing-constant.} \qquad (21)
\end{aligned}$$

We equate the values and gradients of $\log P(A, F)$ and $M(A, F)$ at the touching point (A^i, F^i) to get

$$c^j_{(l,v)} := \frac{q}{2}((a^{j^i}_{(l,v)})^2 + \epsilon)^{\frac{q}{2}-1}; \qquad (22)$$

$$h^j_{(l,v),(\boldsymbol{l}',\boldsymbol{v}')} = 0.5\omega_{(l,v),(\boldsymbol{l}',\boldsymbol{v}')} \text{ when } |a^{j^i}_{(l,v)} - a^{j^i}_{(\boldsymbol{l}',\boldsymbol{v}')}| \leqslant \gamma, \qquad (23)$$

$$h^j_{(l,v),(\boldsymbol{l}',\boldsymbol{v}')} = \frac{\gamma\omega_{(l,v),(\boldsymbol{l}',\boldsymbol{v}')}}{(2((a^{j^i}_{(l,v)} - a^{j^i}_{(\boldsymbol{l}',\boldsymbol{v}')})^2 + \epsilon)^{0.5})} \text{ otherwise,} \qquad (24)$$

and the constants $c'_{(l,v)}$ and $h'^j_{(l,v),(\boldsymbol{l}',\boldsymbol{v}')}$. However, $c'_{(l,v)}$ and $h'^j_{(l,v),(\boldsymbol{l}',\boldsymbol{v}')}$ are absorbed into the log-normalizing-constants, because they are an additive constants, and therefore need *not* be optimized.

REFERENCES

[1] B. Biswal, F. Yetkin, V. Haughton, and J. Hyde, "Functional connectivity in the motor cortex of resting human brain using echo-planar MRI," *Magn. Reson. Med.*, vol. 34, pp. 537–541, 1995.

[2] M. Lee, C. Smyser, and J. Shimony, "Resting-state fMRI: A review of methods and clinical applications," *Amer. J. Neuroradiol.*, vol. 34, no. 10, pp. 1866–1872, 2013.

[3] C. Caballero-Gaudes and R. Reynolds, "Methods for cleaning the BOLD fMRI signal," *NeuroImage*, vol. 154, pp. 128–149, 2017.

[4] V. Renvall and R. Hari, "Transients may occur in functional magnetic resonance imaging without physiological basis," *Proc. Nat. Acad. Sci.*, vol. 106, no. 48, pp. 20510–20514, 2009.

[5] S. Moeller *et al.*, "Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI," *Magn. Reson. Med.*, vol. 63, no. 5, pp. 1144–1153, 2010.

[6] L. Chen *et al.*, "Evaluation of highly accelerated simultaneous multislice EPI for fMRI," *NeuroImage*, vol. 104, pp. 452–459, 2015.

[7] P. Kulkarni, K. Gupta, S. Merchant, and S. Awate, "R-fMRI reconstruction from k-t undersampled simultaneous-multislice (SMS) MRI with controlled aliasing: Towards higher spatial resolution," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2020, pp. 1060–1064.

[8] M. Chiew, S. Smith, P. Koopmans, N. Graedel, T. Blumensath, and K. Miller, "k-t FASTER: Acceleration of functional MRI data acquisition using low rank constraints," *Magn. Reson. Med.*, vol. 74, pp. 353–364, 2014.
[9] L. Chaari, P. Ciuciu, S. Meriaux, and J. Pesquet, "Spatio-temporal wavelet regularization for parallel MRI reconstruction: Application to functional MRI," *Magn. Reson. Mater. Phy.*, vol. 27, pp. 509–529, 2014.
[10] R. Meszlényi, K. Buza, and Z. Vidnyánszky, "Resting state fMRI functional connectivity-based classification using a convolutional neural network architecture," *Front. Neuroinform.*, vol. 11, 2017, Art. no. 61.
[11] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "SENSE: Sensitivity encoding for fast MRI," *Magn. Reson. Med. Official J. Int. Soc. Magn. Reson. Med.*, vol. 42, no. 5, pp. 952–962, 1999.
[12] M. Griswold *et al.*, "Generalized autocalibrating partially parallel acquisitions (GRAPPA)," *Magn. Reson. Med.*, vol. 47, no. 6, pp. 1202–1210, 2002.
[13] P. Roemer, W. Edelstein, C. Hayes, S. Souza, and O. Mueller, "The NMR phased array," *Magn. Reson. Med.*, vol. 16, no. 2, pp. 192–225, 1990.
[14] F. A. Breuer, M. Blaimer, R. M. Heidemann, M. F. Mueller, M. A. Griswold, and P. M. Jakob, "Controlled aliasing in parallel imaging results in higher acceleration (CAIPIRINHA) for multi-slice imaging," *Magn. Reson. Med.*, vol. 53, no. 3, pp. 684–691, 2005.
[15] D. Larkman, J. Hajnal, A. Herlihy, G. Coutts, I. Young, and G. Ehnholm, "Use of multicoil arrays for separation of signal from multiple slices simultaneously excited," *J. Magn. Reson. Imag.*, vol. 13, no. 2, pp. 313–317, 2001.
[16] B. Zahneisen, T. Ernst, and B. Poser, "SENSE and simultaneous multislice imaging," *Magn. Reson. Med.*, vol. 74, no. 5, pp. 1356–1362, 2015.
[17] S. Cauley, J. Polimeni, H. Bhat, L. Wald, and K. Setsompop, "Interslice leakage artifact reduction technique for simultaneous multislice acquisitions," *Magn. Reson. Med.*, vol. 72, no. 1, pp. 93–102, 2014.
[18] K. Setsompop, B. Gagoski, J. Polimeni, T. Witzel, V. Wedeen, and L. Wald, "Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty,"*Magn. Reson. Med.*, vol. 67, no. 5, pp. 1210–1224, 2012.
[19] M. Chiew, N. Graedel, J. McNab, N. Graedel, S. Smith, and K. Miller, "Accelerating functional MRI using fixed-rank approximations and radial-cartesian sampling," *Magn. Reson. Med.*, vol. 76, pp. 1825–1836, 2016.
[20] M. Chiew, N. Graedel, and K. Miller, "Recovering task fMRI signals from highly under-sampled data with low-rank and temporal subspace constraints," *NeuroImage*, vol. 174, pp. 97–110, 2018.
[21] P. Kulkarni, S. Merchant, and S. Awate, "Bayesian reconstruction of R-fMRI from k-t undersampled data using a robust, subject-invariant, spatially-regularized dictionary prior," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2018, pp. 302–306.
[22] Z. Fang, N. Le, M. Choy, and J. Lee, "High spatial resolution compressed sensing (HSPARSE) functional MRI," *Magn. Reson. Med.*, vol. 76, pp. 440–455, 2016.
[23] N. Graedel, J. McNab, M. Chiew, and K. Miller, "Motion correction for functional MRI with three-dimensional hybrid radial-cartesian EPI," *Magn. Reson. Med.*, vol. 78, pp. 527–540, 2017.
[24] P. Aggarwal and A. Gupta, "Accelerated fMRI reconstruction using matrix completion with sparse recovery via split bregman," *Neurocomputing*, vol. 216, pp. 319–330, 2016.
[25] A. Y. Petrov, M. Herbst, and V. Andrew Stenger, "Improving temporal resolution in fMRI using a 3D spiral acquisition and low rank plus sparse (l s) reconstruction," *NeuroImage*, vol. 157, pp. 660–674, 2017.
[26] S. Lingala, Y. Hu, E. DiBella, and M. Jacob, "Accelerated dynamic MRI exploiting sparsity and low-rank structure: K-t SLR," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1042–1054, May 2011.
[27] J. Haldar and Z. Liang, "Spatiotemporal imaging with partially separable functions: A matrix recovery approach," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2010, pp. 716–719.
[28] S. Lingala and M. Jacob, "Blind compressive sensing dynamic MRI," *IEEE Trans. Med. Imag.*, vol. 32, no. 6, pp. 1132–1145, Jun. 2013.
[29] M. de Ridder, K. Klein, and J. Kim, "A review and outlook on visual analytics for uncertainties in functional magnetic resonance imaging," *Brain Inform.*, vol. 5, no. 2, 2018, Art. no. 5.
[30] R. Tanno *et al.*, "Bayesian image quality transfer with CNNs: Exploring uncertainty in dMRI super-resolution," in *Med. Image Comput. Comput. Assist. Intervention 2017*, 2017, pp. 611–619.
[31] J. Bardsley and C. Fox, "An MCMC method for uncertainty quantification in nonnegativity constrained inverse problems," *Inverse Problems Sci. Eng.*, vol. 20, no. 4, pp. 477–498, 2012.
[32] J. Tick, A. Pulkkinen, and T. Tarvainen, "Image reconstruction with uncertainty quantification in photoacoustic tomography," *J. Acoustical Soc. Amer.*, vol. 139, no. 4, pp. 1951–1961, 2016.
[33] J. Fletcher, S. Leng, L. Yu, and C. McCollough, "Dealing with uncertainty in CT images," *Radiology*, vol. 279, no. 1, pp. 5–10, 2016.
[34] J. Gröhl, T. Kirchner, T. Adler, and L. Maier-Hein, "Confidence estimation for machine learning-based quantitative photoacoustics," *J. Imag.*, vol. 4, no. 12, 2018, Art. no. 147.
[35] K. Friston, C. Frith, R. Turner, and R. Frackowiak, "Characterizing evoked hemodynamics with fMRI," *NeuroImage*, vol. 2, no. 2, pp. 157–165, 1995.
[36] G. Salimi-Khorshidi, G. Douaud, C. F. Beckmann, M. F. Glasser, L. Griffanti, and S. M. Smith, "Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers," *NeuroImage*, vol. 90, pp. 449–468, 2014.
[37] S. Jeong, X. Li, J. Yang, Q. Li, and V. Tarokh, "Sparse representation-based denoising for high-resolution brain activation and functional connectivity modeling: A task fMRI study," *IEEE Access*, vol. 8, pp. 36 728–36 740, 2020.
[38] C. Bhushan *et al.*, "Temporal non-local means filtering reveals real-time whole-brain cortical interactions in resting fMRI," *PloS One*, vol. 11, no. 7, pp. 1–22, 2016.
[39] Z. Yang, X. Zhuang, K. Sreenivasan, V. Mishra, T. Curran, and D. Cordes, "A robust deep neural network for denoising task-based fMRI data: An application to working memory and episodic memory," *Med. Image Anal.*, vol. 60, 2020, Art. no. 101622.
[40] P. Kulkarni, S. Merchant, and S. Awate, "R-fMRI reconstruction from k-t undersampled data using a subject-invariant dictionary model and VB-EM with nested minorization," *Med. Imag. Anal.*, vol. 65, 2020, Art. no. 101752.
[41] I. Selesnick, R. Baraniuk, and N. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Process. Mag.*, vol. 22, no. 6, pp. 123–151, Nov. 2005.
[42] M. Novey, T. Adali, and A. Roy, "A complex generalized Gaussian distribution-characterization, generation, and estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1427–1433, Mar. 2010.
[43] E. T. Jaynes, "Information theory and statistical mechanics," *Phys. Rev.*, vol. 106, no. 4, pp. 620–630, 1957.
[44] M. Beal and Z. Ghahramani, "The variational Bayesian EM algorithm for incomplete data: With application to scoring graphical model structures," *Bayesian Statist.*, vol. 7, 2003, pp. 453–464, 2003.
[45] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers,"*FTMI*, vol. 3, no. 1, pp. 1–122, 2011.
[46] M. Lustig, J. Santos, D. Donoho, and J. Pauly, "K-T SPARSE: High frame rate dynamic MRI exploiting spatio-temporal sparsity," in *Proc. 13th Annu. Meeting ISMRM*, 2006.
[47] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Imag. Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[48] M. Welvaert and Y. Rosseel, "On the definition of signal-to-noise ratio and contrast-to-noise ratio for fMRI data," *PloS One*, vol. 8, no. 11, pp. 1–10, 2013.
[49] L. Huang, X. Wang, M. Baliki, L. Wang, A. Apkarian, and T. Parrish, "Reproducibility of structural, resting-state BOLD and DTI data between identical scanners," *PloS One*, vol. 7, 2012, Art. no. e47684.
[50] P. Yoo *et al.*, "7T-fMRI: Faster temporal resolution yields optimal BOLD sensitivity for functional network imaging specifically at high spatial resolution," *NeuroImage*, vol. 164, pp. 214–229, 2018.
[51] I. Drobnjak, D. Gavaghan, E. Suli, J. Pitt-Francis, and M. Jenkinson, "Development of a functional magnetic resonance imaging simulator for modeling realistic rigid-body motion artifacts," *Magn. Reson. Med.*, vol. 56, no. 2, pp. 364–380, 2006.