Signal Processing 157 (2019) 170–179

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Real-time filtering with sparse variations for head motion in magnetic resonance imaging

Daniel S. Weller[a,*], Douglas C. Noll[b], Jeffrey A. Fessler[b]

[a] University of Virginia, Charlottesville, VA 22904, United States
[b] University of Michigan, Ann Arbor, MI 48109, United States

ARTICLE INFO

Article history:
Received 6 July 2018
Revised 22 October 2018
Accepted 3 December 2018
Available online 3 December 2018

Keywords:
Image processing
Registration
Kalman filtering
Sparsity
Magnetic resonance imaging

ABSTRACT

Estimating a time-varying signal, such as head motion from magnetic resonance imaging data, becomes particularly challenging in the face of other temporal dynamics such as functional activation. This paper describes a new Kalman-filter-like framework that includes a sparse residual term in the measurement model. This additional term allows the extended Kalman filter to generate real-time motion estimates suitable for prospective motion correction when such dynamics occur. An iterative augmented Lagrangian algorithm similar to the alterating direction method of multipliers implements the update step for this Kalman filter. This paper evaluates the accuracy and convergence rate of this iterative method for small and large motion in terms of its sensitivity to parameter selection. The included experiment on a simulated functional magnetic resonance imaging acquisition demonstrates that the resulting method improves the maximum Youden's J index of the time series analysis by $2-3\%$ versus retrospective motion correction, while the sensitivity index increases from 4.3 to 5.4 when combining prospective and retrospective correction.

1. Introduction

Prospective head motion correction [1–5] is an example of real-time estimation. Classical approaches to real-time motion tracking frequently involve a linearized (extended or unscented) Kalman filter and assume time series variations follow a Gauss–Markov model. This model efficiently shares information through time, favoring smooth transitions consistent with known (or estimated) autocorrelation matrices for both measurement noise and the random process of innovations influencing state transitions. In magnetic resonance imaging (MRI), PROMO [6] is a widely used example of such a motion correction framework, applying an extended Kalman filter to track a state vector composed of rigid motion (translation and rotation) measured from acquired "navigator data". In addition to improving the statistical reliability of time series analysis in functional magnetic resonance imaging [7], prospective motion correction methods aid volumetric morphometry and other forms of quantitation on very high resolution brain maps [8–10] acquired in vivo. Unlike retrospective methods such as image registration, prospective motion tracking mitigates significant higher-order effects on the acquisition, such as spin history variation caused by motion disrupting the steady state between MRI excitation and relaxation. Accurate and precise motion tracking is necessary in such applications. However, real-time estimation of motion from the raw frequency-domain data ("k-space") used to obtain functional or dynamic MRI's is complicated by overlapping variations from motion and other physiological signals.

1.1. Our contribution

This paper describes a new Kalman-filter-like real-time framework based on a combination of sparse and linearized Gauss–Markov signal and measurement models. While obtaining k-space data, applying this approach facilitates nearly continuous tracking of rigid motion in three dimensions. An appropriate real-time feedback mechanism in an MRI scanner can update the scan prescription over time to follow the estimated motion, mitigating both misalignments and higher-order distortions of the time series signal. We simulate our proposed motion correction and demonstrate improvement in sensitivity and specificity over existing registration corrections when detecting simulated brain activity.

This paper extends our previous work [11] by introducing slice-by-slice motion correction, by studying parameter selection effects on convergence of our iterative estimation algorithm, and by accounting for spin history and other effects encountered in real MRI scanning environments.

* Corresponding author.
E-mail addresses: dsw8c@virginia.edu, dweller@alum.mit.edu (D.S. Weller), dnoll@umich.edu (D.C. Noll), fessler@umich.edu (J.A. Fessler).

https://doi.org/10.1016/j.sigpro.2018.12.001
0165-1684/

1.2. Related work

See Refs. [1,4,12–14] for a comprehensive review of motion correction and image registration in MRI. The linearized motion model used in prospective correction appears in both PACE [15] and PROMO [6]. PACE estimates motion directly from functional images, while PROMO forms Kalman-filter-like estimates from special navigator data interspersed throughout the acquisition. Our technique combines PACE and PROMO, correcting motion directly from functional images, and using a real-time Kalman-filter-like design. Thus, additional navigator sequences are not needed to use our algorithm. Also, our motion correction algorithm improves over PACE by estimating motion continuously, rather than having to wait to obtain a complete volume. To accomplish this, we model temporally sparse changes in the image time series, enabling us to separate motion-related changes from dynamics such as functional activation or blood perfusion. Some retrospective motion correction techniques like k-t FOCUSS [16] and MASTeR [17] employ a similar sparse time residual, but they process the entire time series post-acquisition to suppress non-rigid motion via a block matching approach without any motion model. Unlike those techniques, and other recent sparsity-based reconstruction approaches, the proposed method is designed to correct motion during the acquisition rather than after the fact. As the quality of the motion estimates matters more than the quality of the intermediate images reconstructed in this prospective setting, we do not consider accelerated high-resolution imaging techniques like parallel imaging or nonlinear or data-driven time series modeling approaches.

Another approach would treat the image at each time frame as a dynamic state and rigid motion as unknown parameters of the state transition model. The nonlinear dynamical system would include an extra sparse, non-Gaussian state innovation term to capture unknown temporal dynamics. A linearized Kalman filter can recursively update the motion parameters and the image state estimates, assuming the process variance is fixed [18], even with time delay in the measurements [19] or autoregressive moving average (ARMA) process noise [20]. The time delay is related to slice-by-slice motion estimation, but our approach must handle different time delays for each slice. Beyond Kalman filtering, other examples include nonlinear Wiener filtering [21] and H_∞ filters that constrain worst-case performance [22] and can handle randomly occurring faults, modeled as outliers [23]. Alternatively, after linearizing the effect of motion on the system, the problem of estimating both motion and image states becomes bilinear, leading to maximum likelihood, hierarchical identification, and stochastic gradient methods [24–26]. The proposed framework is distinct from all these approaches, as we do not impose a stochastic model on the image time series, only on the rigid motion parameters.

While this work describes the first addition of a sparse image difference model with a linearized Gauss–Markov model for motion state estimation, the idea of integrating sparsity and Kalman filtering is not new. Some methods replace the Gaussian assumption on the state innovation process with a heavy-tailed sparsity-promoting distribution leading to a 1-norm penalty on the process variation [27,28]. Those methods [29–34] add sparse modeling of the state vector directly, improving estimation quality when prior information about the state's sparsity is known. That approach can be useful for dynamic or functional MRI reconstruction, where the time series images are known to be compressible via a sparsifying transform [35]. While integrating signal or innovation sparsity with the Kalman filter is essentially solved for when the filter's process autocorrelation matrix is known, online estimation of the process autocorrelation matrix becomes more challenging when accounting for sparse innovations. Various applications, such as tracking asset returns in high frequency market data, motivate new covariance estimation approaches [36]. However, none of these articles consider the role of sparse changes in the images being observed, as distinct from the motion parameters being tracked. Furthermore, adapting the process autocorrelation matrix remains unsolved in such settings.

In this paper, the sparse innovations appear in the measurement model instead of the state transition model, so extensions involving non-Gaussian noise and outliers in the measurements are also of interest. Unscented Kalman filters [37] can adapt to measurement uncertainty, but they cannot absorb sudden significant changes from sparse innovations. Adaptive filters also exist for measurement disturbances combining Gaussian noise and outliers, such as those that arise while measuring the state of pneumatic servo actuators [38]. A modified extended Masreliez–Martin filter is proposed for such nonlinear systems [39]. An important result here is that the method is consistent even when the measurements follow a non-Gaussian distribution. For missing measurements, a probabilistic distribution on the missing data enables robust estimation in settings like complex sensor networks with real, physical constraints [40]. While these measurement outlier and missing data models are similar to our sparse innovations model, the sparse innovations due to functional brain activity occur in the image domain, not the measurement domain, requiring a different solution. For instance, Monte Carlo methods can adapt state estimates to abruptly changing environments, given a prior model for the sparse innovations [41]. However, a prior on functional activity in brain imaging may not be available.

Treating the motion as unknown parameters of a nonlinear dynamical system, metaheuristic techniques can mitigate the nonconvexity of the overall estimation problem. For instance, differential evolution evaluates several candidate parameter values in parallel [42]. However, these methods would have to incorporate the unknown sparse innovations in the cost function used to compare solutions to the parameter estimation problem. Related heuristics like parallel particle swarming, grey wolf pack-based optimization, or humpback whale-based spiraling optimization over the entire search space, can attempt to solve for these parameters directly, or in combination with conventional Kalman-based techniques [43,44]. Another such heuristic, the cuckoo search uses a heavy-tailed step length to explore the parameter space more efficiently [45]. The firefly algorithm models the patterns of light flashes used by fireflies to attract mates, and leads to very efficient exploration of this space as well [46]. These metaheuristic methods can jointly estimate state information and optimize the parameters simultaneously, such as for controlling parallel robotic platforms [47]. An improved metaheuristic search inspired by bats can improve the worst-case performance of such platforms, even over the cuckoo search and other common approaches [48]. Hypothetically, such metaheuristic search algorithms can tune other implementation parameters, such as the regularization or penalty parameters in the proposed motion estimation implementation, even though such parameters are not directly part of the state space model.

As a final alternative, artificial neural networks can learn to produce state estimates from a series of observations, incorporating nonlinearity without having to solve a difficult optimization problem in real time (training can be performed off-line if desired). For instance, a neural network derived for battery charge state estimation automatically adapts to the nonlinearity of the state estimation and can accommodate constraints in combination with a particle swarming algorithm [49]. In other work, a generative convolutional neural network can recover the latent dynamical state of nonlinear brain oscillations in magnetoencephalography (MEG) data without being constrained by the tractability of complicated noise models or a limited set of available training data [50]. Such methods can outperform extended or unscented Kalman smoothers in such applications, but the network complexity, and

the amount of training data, would have to grow significantly to accommodate the larger problem size.

1.3. Outline

Section 2 introduces the Kalman filter formulation with sparse signal changes over time and describes a slice-by-slice motion correction framework for the same. Section 3 describes the simulation design, evaluation criteria, and the algorithm parameters used. Section 4 presents experimental results regarding convergence analysis, parameter selection, motion estimation, and time series analysis. Section 5 discusses these results.

2. Theory

Consider complex vector-valued functions of time $\boldsymbol{x}(t)$. Without loss of generality, assume the time t are scaled in units of sampling frames, so the time series consists of vectors $\boldsymbol{x}_0 = \boldsymbol{x}(0)$, $\boldsymbol{x}_1 = \boldsymbol{x}(1)$, and so on. For BOLD functional MRI [51], these vectors represent the brain anatomy with T_2^*-weighted BOLD contrast for each of N_{xy} pixels in N_s slices acquired in a usually interleaved sequence repeatedly over time, so $\boldsymbol{x}_{n,\,t}$ contains the pixel values for the nth slice at time t. The spatial Fourier transform domain k-space data obtained via a Cartesian grid sampling pattern such as echo planar imaging (EPI) are related to these images via the 2D discrete Fourier transform (DFT) $\boldsymbol{F}$:

$$\boldsymbol{d}_{n,t} = \boldsymbol{F}\boldsymbol{x}_{n,t} + \boldsymbol{\xi}_{n,t}, \tag{1}$$

where $\boldsymbol{\xi}_{n,\,t}$ is a vector of iid circularly symmetric complex Gaussian noise with variance σ^2. If k-space is undersampled, then the matrix $\boldsymbol{F}$ would exclude the corresponding rows of the DFT matrix. In general, real MRI acquisitions agree with this noise model, although occasional fluctuations due to radiofrequency interference or obvious spikes at a few individual k-space frequencies do occur. Due to the obviousness of such disturbances in the data, simple thresholding could exclude such disturbances from the DFT matrix used to reconstruct $\boldsymbol{x}_{n,\,t}$.

To describe the evolution of these images, we introduce an auxiliary vector $\boldsymbol{\alpha}_{n,\,t}$ that describes how the corresponding signal relates to the reference frame, $\boldsymbol{x}_{:,\,\mathrm{ref}}$, where the ":" notation indicates the vector contains the entire image stack (not just one slice). In the case of tracking head motion, we use three translation and three rotation parameters to describe the rigid motion throughout the brain. The notation $\boldsymbol{\alpha}^{(-1)}$ describes suitable parameters of the inverse motion transformation, which exists when the motion is rigid. While this is an oversimplification near the neck and base of the cerebellum, it is reasonable in the brain regions that are usually imaged. We describe the process of applying a motion transformation to a volume of stacked slice images as $\boldsymbol{T}(\boldsymbol{\alpha})\boldsymbol{x}_{:,\,t}$. Then, let $\boldsymbol{T}_n(\boldsymbol{\alpha})\boldsymbol{x}_{:,\,t}$ denote extracting the nth slice from the transformed volume. In addition to motion, we allow other dynamics to occur over the course of imaging, such as functional brain activations. These dynamics are reflected in a residual image $\boldsymbol{s}_{n,\,t}$ added to the transformed reference image:

$$\boldsymbol{x}_{n,t} = \boldsymbol{T}_n(\boldsymbol{\alpha}_{n,t})\boldsymbol{x}_{:,\mathrm{ref}} + \boldsymbol{s}_{n,t}. \tag{2}$$

In our slice-by-slice motion correction, slices are acquired at slightly different times, meaning that the motion vectors $\boldsymbol{\alpha}_{1,t}, \boldsymbol{\alpha}_{2,t}, \ldots$ can differ a bit. This possibility even applies to the initial frame, where the first slice is consistent with the reference volume ($\boldsymbol{x}_{1,0} = \boldsymbol{x}_{1,\mathrm{ref}}$), but the others follow the model in Eq. (2) with $\boldsymbol{s}_{n,0} = \boldsymbol{0}$ by definition.

The motion parameter vectors $\boldsymbol{\alpha}_{n,\,t}$ are unknown. To aid estimating the motion, we assume they follow their own state evolution model. A Gauss–Markov model is a reasonable approximation for generating a motion vector $\boldsymbol{\alpha}$ from the previous vector $\boldsymbol{\alpha}_{\mathrm{prev}}$: $p(\boldsymbol{\alpha} \mid \boldsymbol{\alpha}_{\mathrm{prev}}) = \mathcal{N}(\boldsymbol{\alpha}_{\mathrm{prev}}, \boldsymbol{Q})$ with symmetric positive definite covariance matrix $\boldsymbol{Q}$. Indexing the slices in the order they are acquired, $\boldsymbol{\alpha}_{\mathrm{prev}} = \boldsymbol{\alpha}_{n-1,t}$ for $n > 1$, and $\boldsymbol{\alpha}_{N_s,t-1}$ for $n = 1$. By the Markov property, we can relate all the motion parameters $\boldsymbol{\alpha}_{1,t}, \ldots, \boldsymbol{\alpha}_{N_s,t}$ for a time frame to $\boldsymbol{\alpha}_{N_s,t-1}$ via the conditional distribution:

$$p\left(\begin{bmatrix}\boldsymbol{\alpha}_{1,t}\\ \vdots \\ \boldsymbol{\alpha}_{N_s,t}\end{bmatrix} \mid \boldsymbol{\alpha}_{N_s,t-1}\right) = \mathcal{N}\left((\boldsymbol{1}_{N_s\times 1} \otimes \boldsymbol{I})\boldsymbol{\alpha}_{N_s,t-1},\ \boldsymbol{\mathcal{Q}} \triangleq \begin{bmatrix}1 & 1 & \cdots & 1\\ 1 & 2 & \cdots & 2\\ \vdots & \vdots & \ddots & \vdots\\ 1 & 2 & \cdots & N_s\end{bmatrix} \otimes \boldsymbol{Q}\right) \tag{3}$$

This enlarged covariance matrix $\boldsymbol{\mathcal{Q}}$ is symmetric positive definite, and its inverse is a block tridiagonal matrix with closed form (for $N_s > 1$)

$$\boldsymbol{\mathcal{Q}}^{-1} = \begin{bmatrix}2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1\\ & & & -1 & 1\end{bmatrix} \otimes (\boldsymbol{Q}^{-1}). \tag{4}$$

For $N_s = 1$, $\boldsymbol{\mathcal{Q}}^{-1} = \boldsymbol{Q}^{-1}$. Let $\boldsymbol{a}_{n,\,t}$ denote the difference between $\boldsymbol{\alpha}_{n,\,t}$ and $\boldsymbol{\alpha}_{N_s,t-1}$. Combining these models yields the state and measurement equations:

$$\boldsymbol{\alpha}_{n,t} = \boldsymbol{\alpha}_{N_s,t-1} + \boldsymbol{a}_{n,t}, \quad n = 1, \ldots, N_s, \tag{5}$$

$$\boldsymbol{d}_{n,t} = \boldsymbol{F}(\boldsymbol{T}_n(\boldsymbol{\alpha}_{n,t})\boldsymbol{x}_{:,\mathrm{ref}} + \boldsymbol{s}_{n,t}) + \boldsymbol{\xi}_{n,t}, \quad n = 1, \ldots, N_s. \tag{6}$$

A challenge here is that the transformation $\boldsymbol{T}_n(\boldsymbol{\alpha})$ is a nonlinear function of $\boldsymbol{\alpha}$. The sequel describes a linearized Kalman-filter-like implementation to update $\boldsymbol{\alpha}$ as new k-space measurements arrive, facilitating prospective motion correction.

2.1. Slice-by-slice extended Kalman filter

Here, we linearize the measurement model around the present motion estimate. Call $\mathcal{J}\{\boldsymbol{T}_n(\cdot)\boldsymbol{x}\}(\boldsymbol{\alpha}_{n,t})$ the Jacobian matrix of first derivatives with respect to $\boldsymbol{\alpha}$ of the transform $\boldsymbol{T}_n(\boldsymbol{\alpha})\boldsymbol{x}$ for the nth slice, evaluated at $\boldsymbol{\alpha} = \boldsymbol{\alpha}_{n,t}$. For convenience, we write the Jacobian matrix for this slice and time frame as just $\boldsymbol{J}_{n,\,t}$. Using this matrix, the first-order Taylor series expansion of $\boldsymbol{T}_n(\boldsymbol{\alpha})\boldsymbol{x}_{:,\,\mathrm{ref}}$ around $\boldsymbol{\alpha} = \boldsymbol{\alpha}_{N_s,t-1}$ is $\boldsymbol{T}_n(\boldsymbol{\alpha}_{N_s,t-1})\boldsymbol{x}_{:,\mathrm{ref}} + \boldsymbol{J}_{n,t}(\boldsymbol{\alpha} - \boldsymbol{\alpha}_{N_s,t-1}) + o(\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_{N_s,t-1}\|^2)$. When motion does not change too rapidly, the $o(\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_{N_s,t-1}\|^2)$ term decays to zero. The linearized measurement model becomes

$$\boldsymbol{d}_{n,t} = \boldsymbol{F}(\boldsymbol{T}_n(\boldsymbol{\alpha}_{N_s,t-1})\boldsymbol{x}_{:,\mathrm{ref}} + \boldsymbol{J}_{n,t}(\boldsymbol{\alpha}_{n,t} - \boldsymbol{\alpha}_{N_s,t-1}) + \boldsymbol{s}_{n,t}) + \boldsymbol{\xi}_{n,t}. \tag{7}$$

Let $\bar{\boldsymbol{x}}_{n,t-1} = \boldsymbol{T}_n(\boldsymbol{\alpha}_{N_s,t-1})\boldsymbol{x}_{:,\mathrm{ref}}$ with the appropriate estimate of $\boldsymbol{\alpha}_{N_s,t-1}$. Then, define the measurement residual $\boldsymbol{r}_{n,t} = \boldsymbol{d}_{n,t} - \boldsymbol{F}(\bar{\boldsymbol{x}}_{n,t-1} + \boldsymbol{s}_{n,t})$ to be the difference between $\boldsymbol{d}_{n,\,t}$ and its prediction assuming no additional motion.

Since the measurements (and the images) are complex-valued, but the motion parameter vectors are real-valued, we must account for the real $[\,\cdot\,]_R$ and imaginary $[\,\cdot\,]_I$ parts of these variables in the Kalman filter expressions that follow. Given the previous state estimate $\hat{\boldsymbol{\alpha}}_{:,t-1|t-1}$ and previous error covariance estimate $\boldsymbol{P}_{t-1|t-1}$, the Kalman filter prediction step is straightforward: $\hat{\boldsymbol{\alpha}}_{n,t|t-1} = \hat{\boldsymbol{\alpha}}_{N_s,t-1|t-1}$, and $\boldsymbol{P}_{t|t-1} = (\boldsymbol{1}_{N_s\times 1} \otimes \boldsymbol{I})\boldsymbol{P}_{N_s,t-1|t-1}(\boldsymbol{1}_{N_s\times 1}^T \otimes \boldsymbol{I}) + \boldsymbol{\mathcal{Q}}$, where $\boldsymbol{P}_{N_s,t-1|t-1}$ is the part of the error covariance for the motion of the last slice (since predictions for all the slices are with respect to that last slice's value).

To compute the Kalman filter update steps, following the incorporation of new data $\boldsymbol{d}_{:,t}$, we first consider $\boldsymbol{s}_{n,t}$ to be fixed and known. Through some matrix algebra, we have that the motion innovations $\boldsymbol{a}_{:,t}$ for all slices are estimated as

$$\hat{\boldsymbol{a}}_{:,t|t} = \left(\boldsymbol{P}_{t|t-1}^{-1} + \tfrac{1}{\sigma^2}\left[\boldsymbol{J}_{:,t}^H(\boldsymbol{I}_{N_s\times N_s}\otimes \boldsymbol{F}^H\boldsymbol{F})\boldsymbol{J}_{:,t}\right]_R\right)^{-1} \cdot \tfrac{1}{\sigma^2}\left[\boldsymbol{J}_{:,t}^H(\boldsymbol{I}_{N_s\times N_s}\otimes \boldsymbol{F}^H)\boldsymbol{r}_{:,t}\right]_R. \quad (8)$$

Then, $\hat{\boldsymbol{\alpha}}_{:,t|t} = \hat{\boldsymbol{\alpha}}_{:,t|t-1} + \hat{\boldsymbol{a}}_{:,t|t}$. This $\hat{\boldsymbol{a}}_{:,t|t}$ also solves the least squares problem

$$\hat{\boldsymbol{a}}_{:,t|t} = \arg\min_{\boldsymbol{a}} \tfrac{1}{2\sigma^2}\|(\boldsymbol{I}_{N_s\times N_s}\otimes \boldsymbol{F})\boldsymbol{J}_{:,t}\boldsymbol{a} - \boldsymbol{r}_{:,t}\|^2 + \tfrac{1}{2}\boldsymbol{a}^T\boldsymbol{P}_{t|t-1}^{-1}\boldsymbol{a}. \quad (9)$$

Using some more algebra, the posterior error covariance estimate $\boldsymbol{P}_{t|t}$ satisfies

$$\boldsymbol{P}_{t|t}^{-1} = \boldsymbol{P}_{t|t-1}^{-1} + \tfrac{1}{\sigma^2}\left[\boldsymbol{J}_{:,t}^H(\boldsymbol{I}_{N_s\times N_s}\otimes \boldsymbol{F}^H\boldsymbol{F})\boldsymbol{J}_{:,t}\right]_R. \quad (10)$$

When the residual $\boldsymbol{s}_{n,t}$ is not known, we impose a prior model on it. For dynamic MRI, a reasonable model is that the residual is sparse, assuming the dynamic changes not due to motion are localized in space and time. Thus, we expand the optimization to solve for both $\hat{\boldsymbol{a}}_{:,t|t}$ and $\hat{\boldsymbol{s}}_{:,t}$ together:

$$\{\hat{\boldsymbol{a}}_{:,t|t}, \hat{\boldsymbol{s}}_{:,t}\} = \arg\min_{\boldsymbol{a},\boldsymbol{s}} \Psi(\boldsymbol{a},\boldsymbol{s}) \triangleq \tfrac{1}{2\sigma^2}\|(\boldsymbol{I}_{N_s\times N_s}\otimes \boldsymbol{F})(\boldsymbol{J}_{:,t}\boldsymbol{a} + \boldsymbol{s} + \bar{\boldsymbol{x}}_{:,t-1}) - \boldsymbol{d}_{:,t}\|^2 + \tfrac{1}{2}\boldsymbol{a}^T\boldsymbol{P}_{t|t-1}^{-1}\boldsymbol{a} + \lambda\|\boldsymbol{s}\|_1. \quad (11)$$

This convex optimization problem is the key component of our sparse Kalman-filter-like algorithm. While one may solve for $\boldsymbol{a}$ in terms of $\boldsymbol{s}$ and plug back into Eq. (11), the resulting 1-norm-regularized least squares problem involves time-consuming matrix operations. We propose a variable-splitting approach that separates and simplifies the updates for $\boldsymbol{a}$ and $\boldsymbol{s}$. Since the estimate of $\boldsymbol{\alpha}_{:,t}$ is no longer linear, the posterior error covariance estimate $\boldsymbol{P}_{t|t}$ becomes a loose approximation of the true posterior error covariance.

2.2. Initial conditions and covariance estimation

The Kalman filter framework requires a reasonable estimate for $\boldsymbol{\alpha}_{:,-1}$ and $\boldsymbol{P}_{-1|-1}$ consistent with the initial measurement information to perform the initial prediction step. We designate the initial head position as the reference $\boldsymbol{x}_{:,\text{ref}}$, so $\hat{\boldsymbol{\alpha}}_{:,-1} = \boldsymbol{0}$ and $\boldsymbol{P}_{-1|-1} = \boldsymbol{0}$. The initial image $\boldsymbol{x}_{:,\text{ref}}$ can be obtained directly from the initial measurements without considering motion during this time. Thus, the initial conditions are known for this motion estimation problem.

Furthermore, the process and measurement noise covariances $\boldsymbol{Q}$ and $\sigma^2\boldsymbol{I}$ are not necessarily known a priori. In MRI, we can estimate σ^2 by obtaining some additional measurements with the excitation radiofrequency field turned off, However, the process covariance $\mathcal{Q}$ is much harder to access in practice, so we adapt it from a sliding window of motion estimates. Autocovariance least squares (ALS) [52–54] also would work in the linear case. Recent efforts target the cases of nonlinear measurements [55] and sparse innovations [36]. Instead of adopting highly complex methods not necessarily amenable to rapid processing, we constrain $\boldsymbol{Q}$ to be diagonal and form simple estimates from the sample variances of the motion innovations $\hat{\boldsymbol{a}}_{n,t|t}$. This approach ignores cross-correlations across parameters and differences in motion sensitivities for different slices, but prospective correction would dynamically reorient the coordinate system to track the principal direction of the detected motion. So this constraint likely is not a major limitation on the overall algorithm's accuracy.

2.3. AL-based Implementation

Majorization-minimization [56–58] could solve Eq. (11), but guaranteeing a small Lipschitz constant (see supplementary material) for fast convergence to the solution $\{\hat{\boldsymbol{a}}, \hat{\boldsymbol{s}}\}$ is challenging. Instead, we create an auxiliary variable $\boldsymbol{z}$ for the motion-transformed image and use an augmented Lagrangian (AL) scheme similar to the alternating direction method of multipliers (ADMM) [59–62].

First, let $\boldsymbol{z}$ denote the unknown image stack at time t: $\boldsymbol{z} = \bar{\boldsymbol{x}}_{:,t-1} + \boldsymbol{J}_{:,t}\boldsymbol{a} + \boldsymbol{s}$. We define the following modified augmented Lagrangian:

$$\mathcal{L}_A(\boldsymbol{a},\boldsymbol{s},\boldsymbol{z};\boldsymbol{u}) \triangleq \tfrac{1}{2\sigma^2}\|(\boldsymbol{I}\otimes\boldsymbol{F})\boldsymbol{z} - \boldsymbol{d}_{:,t}\|_2^2 + \tfrac{1}{2}\boldsymbol{a}^T\boldsymbol{P}_{t|t-1}^{-1}\boldsymbol{a} + \lambda\|\boldsymbol{s}\|_1 + \tfrac{\mu}{2}\|(\boldsymbol{J}_{:,t}\boldsymbol{a} + \boldsymbol{s} + \bar{\boldsymbol{x}}_{:,t-1}) - \boldsymbol{z} + \boldsymbol{u}\|_2^2, \quad (12)$$

where $\boldsymbol{u}$ and μ are the scaled Lagrange dual vector and augmented Lagrangian penalty parameter, respectively.

Alternating minimizing $\boldsymbol{a}$, $\boldsymbol{s}$, and $\boldsymbol{z}$ would correspond to ADMM. Instead, we update $\boldsymbol{a}$ and $\boldsymbol{z}$ jointly, and separately from $\boldsymbol{s}$:

$$\boldsymbol{s}^{i+1} \leftarrow \arg\min_{\boldsymbol{s}} \lambda\|\boldsymbol{s}\|_1 + \tfrac{\mu}{2}\|(\boldsymbol{J}_{:,t}\boldsymbol{a}^i + \boldsymbol{s}) - \boldsymbol{z}^i + \bar{\boldsymbol{x}}_{:,t-1} + \boldsymbol{u}^i\|_2^2, \quad (13)$$

$$\{\boldsymbol{a}^{i+1}, \boldsymbol{z}^{i+1}\} \leftarrow \arg\min_{\boldsymbol{a},\boldsymbol{z}} \tfrac{1}{2\sigma^2}\|(\boldsymbol{I}\otimes\boldsymbol{F})\boldsymbol{z} - \boldsymbol{d}_{:,t}\|_2^2 + \tfrac{1}{2}\boldsymbol{a}^T\boldsymbol{P}_{t|t-1}^{-1}\boldsymbol{a} + \tfrac{\mu}{2}\|(\boldsymbol{J}_{:,t}\boldsymbol{a} + \boldsymbol{s}^{i+1}) - \boldsymbol{z} + \bar{\boldsymbol{x}}_{:,t-1} + \boldsymbol{u}^i\|_2^2, \quad (14)$$

$$\boldsymbol{u}^{i+1} \leftarrow \boldsymbol{u}^i + (\boldsymbol{J}_{:,t}\boldsymbol{a}^{i+1} + \boldsymbol{s}^{i+1}) - \boldsymbol{z}^{i+1} + \bar{\boldsymbol{x}}_{:,t-1}. \quad (15)$$

Updating $\boldsymbol{s}$ soft-thresholds $\boldsymbol{z}^i - \bar{\boldsymbol{x}}_{:,t-1} - \boldsymbol{J}_{:,t}\boldsymbol{a}^i - \boldsymbol{u}^i$, which elementwise shrinks the complex-valued argument (call it s_n) according to the threshold $\frac{\lambda}{\mu}$: $s_n \leftarrow \text{sign}(s_n)\cdot\max\{0, |s_n| - \frac{\lambda}{\mu}\}$. The normal equations from Eq. (14) yield $\boldsymbol{a}$ and $\boldsymbol{z}$:

$$\begin{aligned}(\boldsymbol{P}_{t|t-1}^{-1} + \mu\left[\boldsymbol{J}_{:,t}^H\boldsymbol{J}_{:,t}\right]_R)\boldsymbol{a} - \mu\left[\boldsymbol{J}_{:,t}^H\boldsymbol{z}\right]_R &= -\mu\left[\boldsymbol{J}_{:,t}^H(\bar{\boldsymbol{x}}_{:,t-1} + \boldsymbol{u}^i + \boldsymbol{s}^{i+1})\right]_R,\\ -\mu\boldsymbol{J}_{:,t}\boldsymbol{a} + \left[\frac{1}{\sigma^2}(\boldsymbol{I}\otimes(\boldsymbol{F}^H\boldsymbol{F})) + \mu\boldsymbol{I}\right]\boldsymbol{z} &= \frac{1}{\sigma^2}(\boldsymbol{I}\otimes\boldsymbol{F}^H)\boldsymbol{d}_{:,t}\\ &\quad +\mu(\bar{\boldsymbol{x}}_{:,t-1} + \boldsymbol{u}^i + \boldsymbol{s}^{i+1}). \end{aligned} \quad (16)$$

Schur complements solve this system of equations efficiently (see supplementary material). The final step of our algorithm updates $\boldsymbol{u}$ via vector addition.

As the problem scales in the image size N_{xy}, number of slices N_s, or number of frames, the update steps of our ADMM-like implementation remain reasonable computationally. As N_{xy} increases, the DFT operation grows like $N_{xy}\log N_{xy}$, and all the other operations involved in solving the normal equations in Eq. (16) grow linearly. Similarly, the matrix-vector inversions scale linearly with N_s, since the matrices involved are either block diagonal or block tridiagonal. Matrix-vector products involving $\boldsymbol{I}\otimes\boldsymbol{F}$ or $\boldsymbol{J}_{:,t}$ also scale linearly in N_s, as they are either stacked or block diagonal matrices. Finally, the computational and storage complexity remains constant as the number of time frames grows, as opposed to methods that process the entire time series jointly.

2.4. Initialization and parameter selection

The convergence rate of the ADMM-like method proposed here is affected by both the initialization and the choice of penalty parameter μ. We initialize $\boldsymbol{a}^0 = \boldsymbol{0}$ by setting the motion for the current slice is the same as the previous one. To initialize $\boldsymbol{z}$ and $\boldsymbol{s}$, we perform a direct reconstruction of the data $\boldsymbol{d}_{:,t}$ to form $\boldsymbol{z}^0$, and assign the residual image $\boldsymbol{z}^0 - \bar{\boldsymbol{x}}_{:,t-1} - \boldsymbol{J}_{:,t}\boldsymbol{a}^0$ to $\boldsymbol{s}^0$.

The choice of penalty parameter μ that leads to fastest possible convergence is expected to vary with the noise and process innovation covariances as well as the choice of λ. AL methods like ADMM converge, even with finite μ [61]. The parameter selection experiment explores how a fixed value of μ can affect convergence for a range of motions. Heuristics for fixing μ based on the numerical conditioning of the quadratic subproblem and the sparse thresholding fraction of the shrinkage subproblem can achieve consistent results across reconstructions [63]. At the same time, adaptive heuristic methods [62] allow μ to vary to balance minimizing the objective function and preserving the equality constraint ($\boldsymbol{z} = \bar{\boldsymbol{x}}_{:,t-1} + \boldsymbol{J}_{:,t}\boldsymbol{a} + \boldsymbol{s}$). Optimizing the choice of μ is also feasible for quadratic problems [64]; this method could approximate non-quadratic problems as well. Thus, we can select fixed or adaptive μ automatically.

The other tuning parameter introduced in our implementation, λ, controls the sparsity of the image innovations $\boldsymbol{s}$, so λ can be considered an implicitly described parameter of the state space model. We propose using a fixed λ during prospective motion correction and adapt the process covariance matrix to abrupt motion changes in real time. Alternatively, automatic parameter selection methods like Stein's unbiased risk estimator [65–67] can help pick λ. Derivative-free metaheuristic search algorithms described in Section 1 can optimize λ to make the residual energy in the data consistency term consistent with the expected noise variance of the data (consistent with the discrepancy principle [68]). However, tuning λ during the motion estimation process would involve calculating and comparing many motion estimates in parallel, so parameter selection ideally would be performed only during the initial stimulus phases.

As an alternative to adjusting λ for large motion, we temporarily scale our estimate of the process innovation covariance $\boldsymbol{Q}$ by a large number whenever we detect large impulsive motion. Doing so temporarily decouples the last motion estimate from the current one, permitting rapid adjustment to the large motion. To detect this large motion, we use a threshold on the total energy of the latter half of slices $\boldsymbol{s}$, since those slices are most affected by the new motion. When the motion is consistent with the random walk model, so the change in motion is well-modeled by the extended Kalman filter, the sparse innovations $\boldsymbol{s}$ for those slices should include only the functional activity and some small residual from matching the measurement noise. The 1-norm sparsity penalty suppresses the latter. Both of these are on the order of a few percent of the total signal energy, so a much larger change can only be explained by a failure of the motion model. Thus, a relatively conservative threshold would be a reliable change detector, and a relatively large scaling of $\boldsymbol{Q}$ would effectively reset the Kalman filter to handle large motion. The process covariance estimation then naturally decreases $\boldsymbol{Q}$ over time as the motion returns to being consistent with the model.

2.5. Sensing matrix for echo planar imaging

Echo planar imaging (EPI) is used often for functional MRI, and is highly susceptible to systematic k-space offsets caused by eddy current effects. These produce aliasing artifacts (ghosts) offset by $N_y/2$ in the phase encode direction. These ghosts disrupt motion estimation, making ghost correction necessary in our reconstruction. We estimate the in-plane k-space shifts through a calibration pre-scan [69,70]. Since prospective motion correction rotates the acquisition readout gradient axis, correcting phase ghosts requires calculating oblique-plane k-space offsets [71]. Shifting the odd and even phase encode lines of our data corrects the ghost artifacts before performing motion estimation. When we include the EPI phase shifts in our forward model, we can still implement the sensing matrix $\boldsymbol{F}$ efficiently using the fast Fourier transform by treating the odd and even EPI lines as two half transforms.

3. Materials and methods

To analyze the proposed Kalman-filter-like framework incorporating a sparse residual model in the images, we implemented our method for prospective motion estimation in MATLAB (The Mathworks, Natick, MA), using the Michigan Image Reconstruction Toolbox, which is available online from http://web.eecs.umich.edu/~fessler/code/index.html, along with MEX files previously developed for B-spline-based interpolation [72,73]. We first evaluate the effect of varying the sparse regularization parameter λ and including $\mathcal{Q}$-adaptation. Then, we vary the AL penalty parameter μ and measure how the objective function convergence changes. Our real-time experiments simulate a realistic BOLD functional MRI acquisition described next.

3.1. Simulated data

Our simulation of BOLD functional MR imaging uses a high-resolution T_2^*-weighted Brainweb phantom [74], shown in Fig. 1 with active regions highlighted. This phantom contains $181 \times 217 \times 181$ voxels, each 1 mm isotropic resolution, with no noise or inhomogeneity effects. From this volume we extract twenty slices, with a uniform slice profile 3-mm thick, covering a 6 cm slab with no gaps. We also construct a high-resolution T_1 map of this phantom for modeling spin history effects, by tracking tissue-specific longitudinal relaxations, corresponding to the time between successive excitations, of each high-resolution voxel. Our simulated acquisition includes spatial variations in intensity caused by different longitudinal relaxations, or "spin histories." We interleave the slice order in "bit-reversed" fashion, with greater distance between the initial slices of each frame, to facilitate three-dimensional motion estimation. We simulate a single-shot EPI acquisition, including small phase variations reflecting the presence of eddy current effects. To account for partial volume effects, k-space samples from this high-resolution phantom are reconstructed at a lower resolution, measuring 4×4 mm in-plane.

To this brain phantom, we add five 3D ellipsoidal regions of varying dimensions to represent sets of activated voxels. All of the activated voxels have added to their amplitudes an activation time series (shown in Fig. 2) corresponding to a block design task repeated three times over 200 s (200 frames with TR=1 s), with 30 s off, and 30 s on. This block design is convolved with a canonical hemodynamic response function [75] to emulate the vascular response to activation. These activations produce signal changes of approximately 3% of the maximum image intensity.

We simulate motion for each slice acquisition in every frame using a combination of two motion models, producing variability not expected by our Kalman filter. The first is the random-walk Markov model discussed in Section 2, with zero-mean Gaussian innovations with a standard deviation of 0.05 mm or degrees, per second (unknown to our motion estimation system). We add to this random walk impulsive motion with a larger magnitude of 1 mm or degree over a second; impulse times are exponentially distributed with a mean occurrence time of 50 s. Each of the six motion parameters are simulated independently. The generated motion trajectories had no more than ± 5 mm or degrees of motion in each direction. When simulating prospective correction, these simulated motions are applied after adjusting the volume coordinates for the motion estimated for the last slice of the previous time frame ($\hat{\boldsymbol{\alpha}}_{N_s,t-1}$).

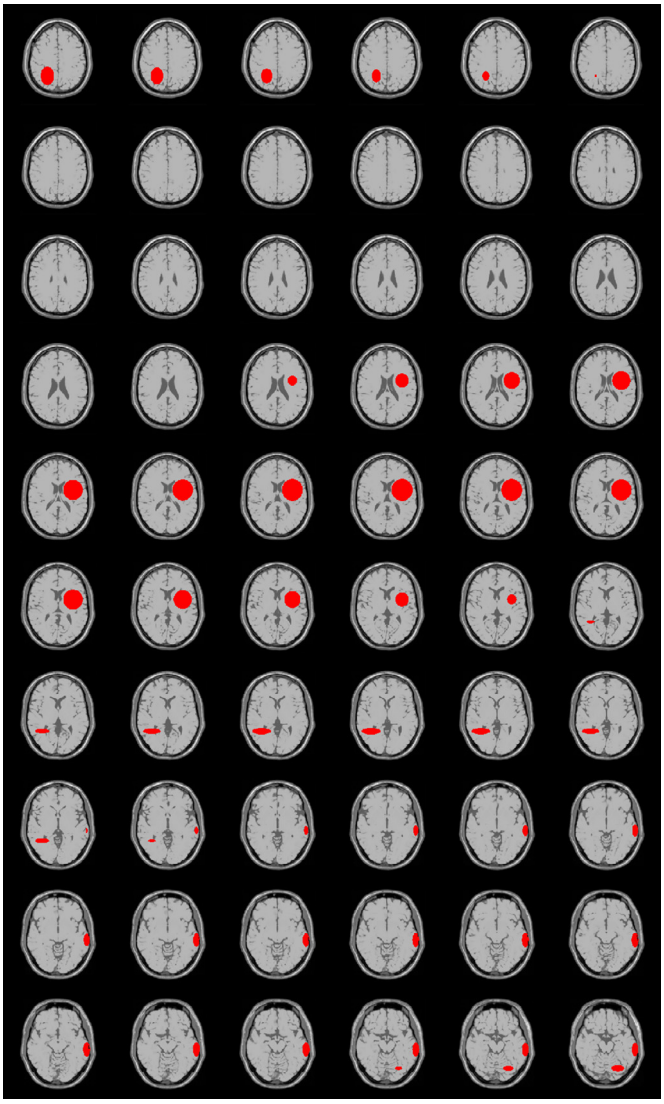

Fig. 1. Active regions (red) are overlaid on the high-resolution T_2^*-weighted Brainweb phantom used in our simulations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

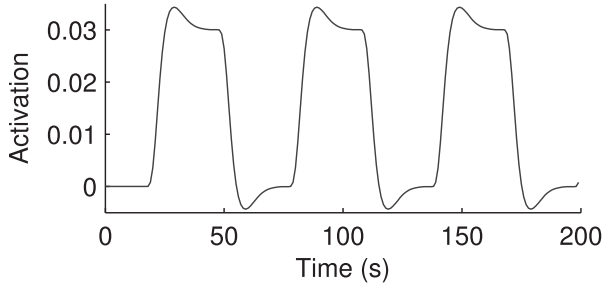

Fig. 2. The activation time series is typical for a block design with alternating off- and on-task blocks.

3.2. Parameter selection

Before running our complete simulation, we study the effects of choosing λ and μ on the performance of our motion estimation. Recognizing that activation patterns may change between scans, and desiring to determine broadly applicable choices of parameters, we exclude functional activations from the parameter selection tests. For simplicity, we also ignore spin history, EPI phase ghosts, quantization, and other effects for the parameter selection experiments. We run a short time series based on the acquisition strategy with realistic slice-by-slice motion for $\lambda \in \{1, 10, 100, 1000\}$. For each λ, we run our method for a range of μ's, retaining the result that achieved the lowest objective function value. To test our adaptive $\boldsymbol{Q}$ scheme, we scale $\boldsymbol{Q}$ by a factor of ten whenever the latter half of slices of $\boldsymbol{s}$ on average contain greater than unit energy (our threshold). We aim to show that adapting $\boldsymbol{Q}$ (with λ held fixed) can provide some of the same advantages of adapting λ in the presence of large, impulsive motion.

Next, we vary the penalty parameter μ to determine a value that yields rapid convergence with both small and large motion. For large impulsive motions, using the adaptive scheme just described would increase $\boldsymbol{Q}$ by several orders of magnitude, yielding much smaller precision matrices $\boldsymbol{P}_{t|t-1}^{-1}$. However, we hold μ fixed to simplify real-time implementation. We run our AL-based method for $\mu \in \{1, 10, 100, 1000, 10^4\}$ and compare the objective function convergence against each other. We aim to show that even a fixed $\mu = 100$ can achieve relatively consistent convergence in both settings (small and large motion).

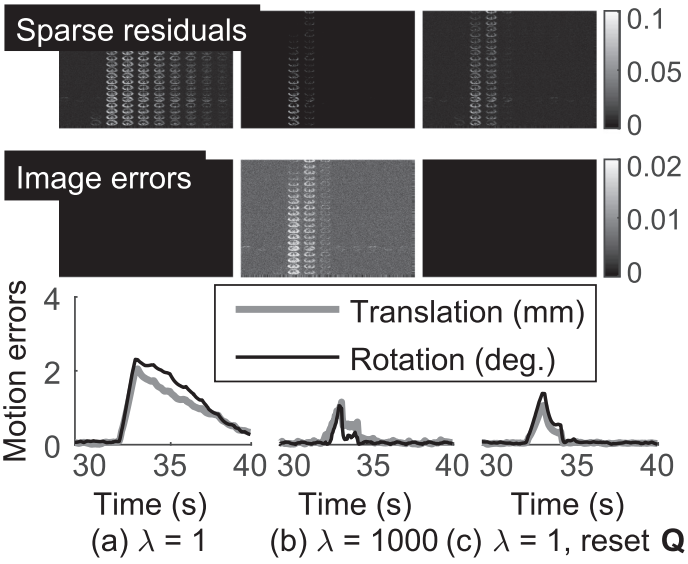

Fig. 3. The sparse residual (top), image error (middle), and motion error (bottom) are plotted for both (a) $\lambda = 1$ and (b) $\lambda = 1000$ with $\boldsymbol{Q}$ held fixed (at the true value for the random walk model). The sparse residual, image error, and motion error also are plotted for (c) $\lambda = 1$, resetting $\boldsymbol{Q}$ when the sparse residual energy is large, such as after the large motion around 33 s.

4. Experimental results

4.1. Parameter selection

Fig. 3 illustrates the effects of increasing λ when large motion, such as from an impulse, is present. The predicted error covariance $\boldsymbol{P}_{t|t-1}$ attempts to keep $\boldsymbol{a}$, the change in motion, small to be consistent with the random walk model. The error due to underestimating the motion appears in the sparse residual $\boldsymbol{s}$. As λ increases, more of this error appears in the image estimate $\boldsymbol{x}$ instead of in the sparse residual, but the translational and rotational errors in the estimated motion decrease much more rapidly. Fig. 3(c) shows that temporarily increasing $\boldsymbol{Q}$ achieves the desired effect of reducing the magnitude and duration of errors due to large residual motion without altering λ. Also, this strategy does not introduce additional errors into the image estimates, unlike increasing λ.

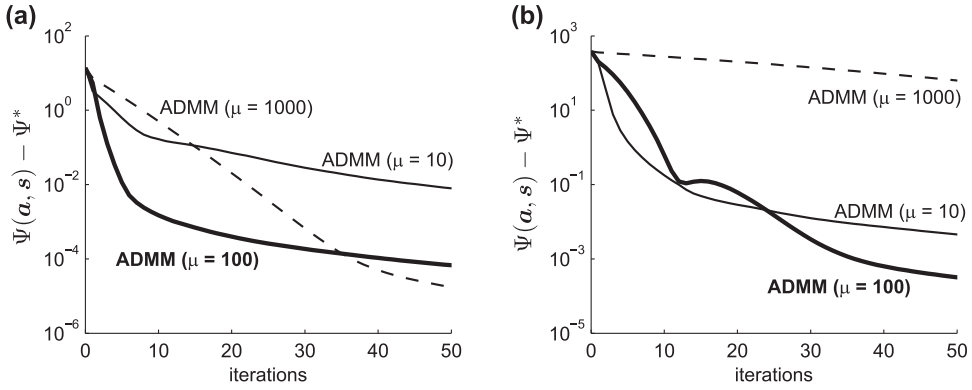

Fig. 4. The objective function $\Psi(\boldsymbol{a}, \boldsymbol{s})$ convergence to a minimum value Ψ^* is plotted for our ADMM-like method with $\mu = \{10, 100, 1000\}$, for both small motion (a) and large motion (b) cases. In both cases, our algorithm rapidly reaches the optimal objective function value over a wide range of μ. Despite small motion and large motion cases using significantly different values of $\mathbf{Q}$ (according to the adaptive scheme), choosing $\mu = 100$ yields good convergence over the first 50 iterations for both cases.

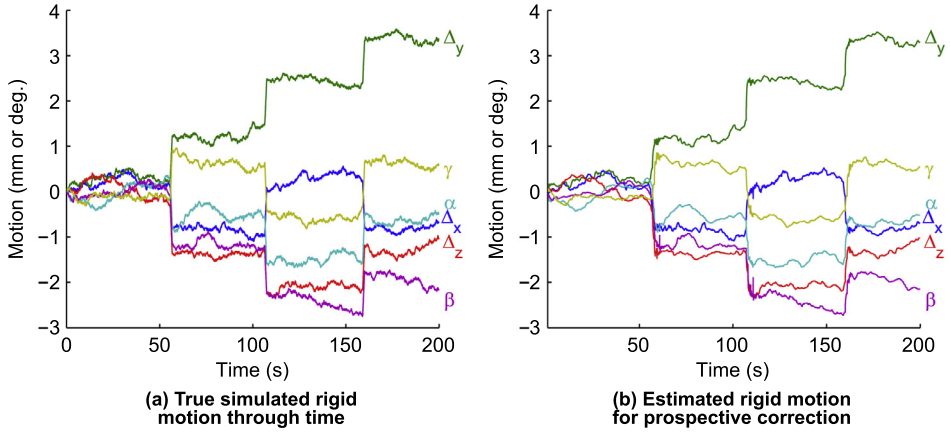

Fig. 5. Ground truth (a) and estimated (b) translations Δ_x, Δ_y, Δ_z and rotations α, β, and γ (relative to the z-, y-, and x-axes, respectively) using the proposed sparse-residual Kalman filter estimation technique.

Fig. 4 plots the objective function $\Psi(\boldsymbol{a}, \boldsymbol{s})$, relative to the long-term-optimal value Ψ^*, for our ADMM-like method with the three fastest-converging values of μ. We depict two cases: frame #20, containing only random walk motion, and frame #34, featuring the aftereffects of large impulsive motion. In both cases, the ADMM-like method converges rapidly in objective function value over a wide range of μ, and the AL penalty parameter $\mu = 100$ performs reasonably in both motion settings. Since the computation time is about the same throughout (9 ms/iteration each), plots of objective function convergence versus time would appear similar to those in Fig. 4.

4.2. Real-time simulated experiments

Our real-time experiments jointly simulate BOLD functional MRI acquisitions and prospective motion estimation on two separate MATLAB instances communicating with each other in real time. Fig. 5 shows that the motion estimates generated by the slice-by-slice prospective correction algorithm, appear very similar to the true motion. The absolute errors for the translational and rotational motion estimates measure 0.063 ± 0.10 mm and 0.085 ± 0.17 degrees, respectively (mean $\pm$ std. dev.). These translational errors are computed as the 2-norm of the Δ_x, Δ_y, and Δ_z errors together. The rotational motion errors are computed as the magnitude of the angle in the axis-angle representation of the difference between the true and estimated rotations, where rotations are relative to the center of the 6 cm slab.

At least as important as the accuracy of the motion estimates is the impact of prospective correction using these measurements. Even if these motion estimates are exact, we expect some residual motion to remain since we apply the previous frame's last motion estimate to acquire the current frame. In prospective correction, the residual motion should remain small relative to the overall motion throughout the course of the scan, so effects like spin history should be minimized. In fact, the average residual translation and rotation decreases from 2.2 ± 1.3 mm to 0.11 ± 0.19 mm, and from 1.6 ± 1.0 degrees to 0.13 ± 0.25 degrees, after prospective correction.

The functional MRI analysis that follows compares prospective, retrospective (the standard realign-and-reslice tool in SPM12 available online at http://www.fil.ion.ucl.ac.uk/spm/), and both motion corrections against an ideal acquisition without head motion. The time series data for each voxel in the reconstructed/realigned volume are then correlated against the true activation time series in Fig. 2. Fig. 6(a)–(d) features slices #9 and #11 with overlaid correlation maps for no motion, retrospective correction, prospective correction, and prospective and retrospective correction combined. These overlays are all thresholded according to a significance threshold of $\alpha = 0.01$, adjusted for multiple comparisons using Bonferroni correction by the number of voxels in the brain. Not shown, the motion-corrupted correlation map without any correction has very low quality. Fig. 6(e) quantitatively compares these fMRI time series analyses over all twenty slices using receiver operating characteristic (ROC) curves to track the sensitivity and specificity of each method for detecting functional activations. The ground truth is generated by reducing the resolution of the high-resolution activation mask shown in Fig. 1. We focus on the low

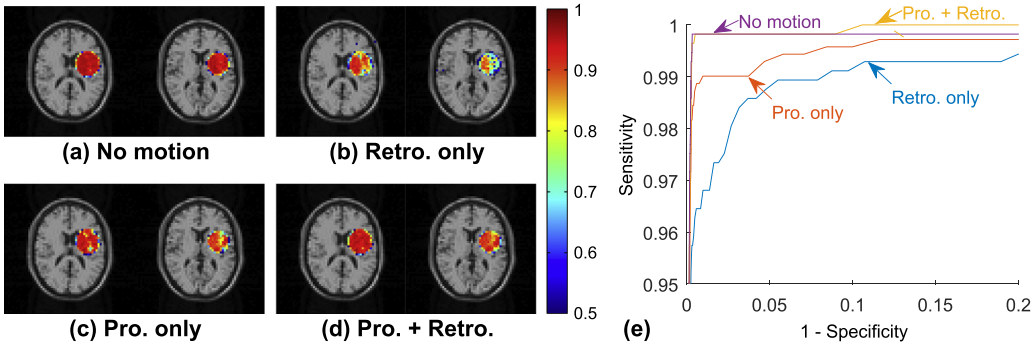

Fig. 6. Correlation maps ($\alpha = 0.01$) are overlaid on the low-resolution reconstructed slices #9 and #11 for (a) no motion (ideal), (b) retrospective correction (using SPM8) only, (c) slice-by-slice prospective correction only, and (d) both prospective and retrospective corrections. The receiver operating characteristic (ROC) curves (e) plot sensitivity versus false alarm rate (1-specificity) for these methods is shown.

Table 1
ROC curve statistics for functional MRI correlation analysis.

	AUC	$\Delta p'$	d'
No correction	0.945	0.81	2.8
Retrospective only	0.994	0.96	4.3
Prospective only	**0.997**	**0.98**	**4.8**
Prospective + retrospective	**> 0.999**	**0.99**	**5.4**
No motion (ideal)	0.998	> 0.99	5.7

false-alarm-rate regime, as activations are usually rare events in task-based functional MRI studies. The area under the curve (AUC) is one summary statistic for quantifying the overall quality of the detection algorithm, but the AUC equally weights all prior probabilities for the true hypothesis, ignoring the scarcity of true activations. Alternatively, another statistical measure of the quality of detection is the maximum Youden's J index $\Delta p'$, which is the maximum difference between sensitivity and false alarm rate. This criterion is related to the sensitivity index d' according to the formula $d' = \Phi^{-1}(\text{sensitivity}) - \Phi^{-1}(1 - \text{specificity})$ for the J-index-maximizing sensitivity and specificity, where $\Phi^{-1}(\cdot)$ is the inverse cumulative density function of the standard Normal distribution. Table 1 lists the AUC, $\Delta p'$, and d' statistics for these ROC curves.

5. Discussion

Ensuring a reasonable convergence rate for this motion estimation method is essential for a nearly real-time application like prospective motion correction, but the fastest possible ADMM convergence is likely not necessary given the progressive nature of prospective motion correction permits the algorithm to make additional corrections to motion through time. As long as retrospective correction after the prospectively corrected acquisition demonstrates that the residual motion errors are roughly of the same scale as the motion changes from frame to frame, the benefits of further improvements in prospective motion estimation accuracy become negligible. Given the motion errors in Fig. 3 are relatively small before and after the large motion, in both the cases of $\lambda = 1000$ and of $\lambda = 1$ with adaptive $\boldsymbol{Q}$, this condition appears to be satisfied by our method with a simple, fixed μ. That said, the adaptive heuristics described could further minimize the number of iterations between accurate motion parameter updates, helping the prospective estimation method to keep up with head motion during a more challenging scan. In our experiments, we keep the number of algorithm iterations between updates below 50 steps. Ensuring our ADMM-like optimization makes sufficient progress towards the optimum during this time is important. Observing in Fig. 4 rapid convergence within this many iterations for both small and big motions confirms the suitability of our optimization method for this simulation.

While we observe that the objective converges quickly using $\mu = 100$ regardless of whether the motion is small (with a realistic covariance $\boldsymbol{Q}$), or large (temporarily increasing the covariance $\boldsymbol{Q}$ by two or three orders of magnitude, to enable larger jumps), further study is needed to demonstrate the broad applicability of $\lambda = 1$, $\mu = 100$ across many data sets. Metaheuristic methods and other automatic parameter selection techniques described earlier may provide a means to tune λ during the earliest task activation, on a per-scan basis. This approach would automate selection of λ without overburdening the real-time estimation after the first task activation period is done (a functional MRI scan frequently runs for many such periods in succession). The sparse residual energy also appears to be a suitable criterion for detecting large impulsive motion, although other techniques like parallel imaging motion detection [76] may also be effective alternatives. Also, observing that $\lambda = 1$ generalizes from the no-activation experiment to our simulation with functional activations supports the practicality of using a fixed parameter without additional tuning.

Significant differences are apparent in the activation correlation maps as a result of using this motion correction method prospectively. The content of these ellipses are noticeably degraded with retrospective correction, likely due to the inability to mitigate spin history effects and the smoothing nature of retrospective interpolation. Prospective correction noticeably improves the uniformity of these ellipses. Prospective correction, either alone or in combination with retrospective correction, also improves the maximum Youden's J index (sensitivity and specificity) of detecting functional activations across a wide range of thresholds, especially in the small false positive regime, the important region for rare events like activations. Combining prospective and retrospective correction provides even greater improvement.

However, our conclusions come with several caveats. Our simulations cannot exactly replicate what we would observe with a real acquisition. Real functional MRI acquisitions are not explicitly task-specific, and data often contain unrelated activations. Other motion-related changes, such as to the magnetic field sensitivities of the receiver antennas used, are not modeled by our method. Our method also does not explicitly track and account for outliers in the measured k-space, and undersampling k-space to accelerate the acquisition also may make motion estimation less accurate. In general, our method assumes our measurements are Gaussian, which may not fit all practical applications. Non-Gaussian noise likely would produce non-sparse errors in the image domain, so a sparse regularizer (such as that applied to $\boldsymbol{s}$) would likely mitigate the effects of such noise or outliers. Furthermore, rigid motion estimation is low-dimensional on its own, so the effects of a

few outliers may be small. Additional evaluation is necessary before applying this method to such environments where such noises are prevalent. Additionally, we do not incorporate time-varying physiological signals that may alias onto the frequencies of our task-related activations during a typical study. Other time-varying effects include slow temperature drift and magnetic field inhomogeneity changes, and these effects may add distortions that do not fit the sparse residual image model. Techniques like bandpass filtering and source separation (e.g., RETROICOR [77]) effectively mitigate some of these perturbations and can be combined with our phase-corrected EPI reconstruction, assuming nearly real-time implementations are available.

6. Conclusion

In conclusion, we developed a novel Kalman-filter-like real-time estimation framework incorporating a sparse residual term in the measurement model to accommodate other time-varying dynamics. We presented two implementations of this framework for prospective motion correction. Our experiments analyzed the tuning parameters for these methods and compared their convergence rates for different degrees of motion. We also demonstrated the effectiveness of this prospective motion correction on a time series analysis of a simulated functional MRI acquisition. We discussed our results and limitations of our experimental setup. In summary, our prospective motion correction method has the potential to significantly improve the sensitivity and specificity of functional MRI time series analysis versus existing retrospective methods.

Declarations of interest

None.

Acknowledgments

Funding: This work was supported by the National Institutes of Health [grant numbers F32EB015914, R21EB022309, and R01EB023618] and the Thomas F. and Kate Miller Jeffress Memorial Trust, Bank of America, Trustee.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.sigpro.2018.12.001.

References

[1] J. Maclaren, M. Herbst, O. Speck, M. Zaitsev, Prospective motion correction in brain imaging: a review, Magn. Reson. Med. 69 (3) (2013). 621–36
[2] J.D. Power, B.L. Schlaggar, S.E. Petersen, Recent progress and outstanding issues in motion correction in resting state fMRI, NeuroImage 105 (2015) 536–551.
[3] M. Goto, O. Abe, T. Miyati, H. Yamasue, T. Gomi, T. Takeda, Head motion and correction methods in resting-state functional MRI, Magn. Reson. Med. Sci. 15 (2) (2016) 178–186.
[4] F. Godenschweger, U. Kägebein, D. Stucht, U. Yarach, A. Sciarra, R. Yakupov, F. Lüsebrink, P. Schulze, O. Speck, Motion correction in MRI of the brain, Phys. Med. Biol. 61 (5) (2016). R32
[5] M. Zaitsev, B. Akin, P. LeVan, B.R. Knowles, Prospective motion correction in functional MRI, NeuroImage 154 (2017) 33–42.
[6] N. White, C. Roddey, A. Shankaranarayanan, E. Han, D. Rettmann, J. Santos, J. Kuperman, A. Dale, PROMO: real-time prospective motion correction in MRI using image-based tracking, Magn. Reson. Med. 63 (1) (2010) 91–105.
[7] K.J. Friston, S. Williams, R. Howard, R.S.J. Frackowiak, R. Turner, Movement-related effects in fMRI time-series, Magn. Reson. Med. 35 (3) (1996). 346–55
[8] M.F. Callaghan, O. Josephs, M. Herbst, M. Zaitsev, N. Todd, N. Weiskopf, An evaluation of prospective motion correction (PMC) for high resolution quantitative MRI, Front. Neurosci. 9 (2015) 97.
[9] D. Stucht, K.A. Danishad, P. Schulze, F. Godenschweger, M. Zaitsev, O. Speck, Highest resolution in vivo human brain MRI using prospective motion correction, PLoS ONE 10 (7) (2015) e0133921.
[10] C. Federau, D. Gallichan, Motion-correction enabled ultra-high resolution in–vivo 7T-MRI of the brain, PLoS ONE 11 (5) (2016) e0154974.
[11] D.S. Weller, J.A. Fessler, Prospective motion correction for functional MRI using sparsity and Kalman filtering, in: Proc. SPIE 8858 Wavelets and Sparsity XV, 2013, p. 885823.
[12] M. Zaitsev, J. Maclaren, M. Herbst, Motion artifacts in MRI: a complex problem with many partial solutions, J. Magn. Reson. Imaging 42 (4) (2015) 887–901.
[13] J.A. Schnabel, M.P. Heinrich, B. W.Papież, S.J.M. Brady, Advances and challenges in deformable image registration: from image fusion to complex motion modelling, Med. Image Anal. 33 (2016) 145–148.
[14] M.A. Viergever, J.A. Maintz, S. Klein, K. Murphy, M. Staring, J.P. Pluim, A survey of medical image registration - under review, Med. Image Anal. 33 (2016) 140–144.
[15] S. Thesen, O. Heid, E. Mueller, L.R. Schad, Prospective acquisition correction for head motion with image-based tracking for real-time fMRI, Magn. Reson. Med. 44 (3) (2000). 457–65
[16] H. Jung, K. Sung, K.S. Nayak, E.Y. Kim, J.C. Ye, K-t FOCUSS: a general compressed sensing framework for high resolution dynamic MRI, Magn. Reson. Med. 61 (1) (2009). 103–16
[17] M.S. Asif, L. Hamilton, M. Brummer, J. Romberg, Motion-adaptive spatio-temporal regularization for accelerated dynamic MRI, Magn. Reson. Med. 70 (3) (2013). 800–12
[18] F. Ding, X. Wang, Q. Chen, Y. Xiao, Recursive least squares parameter estimation for a class of output nonlinear systems based on the model decomposition, Circuits Syst. Signal Process. 35 (9) (2016) 3323–3338.
[19] F. Ding, X. Wang, L. Mao, L. Xu, Joint state and multi-innovation parameter estimation for time-delay linear systems and its convergence based on the Kalman filtering, Digital Signal Process. 62 (2017) 211–223.
[20] F. Ding, D. Meng, J. Dai, Q. Li, A. Alsaedi, T. Hayat, Least squares based iterative parameter estimation algorithm for stochastic dynamical systems with arma noise using the model equivalence, Int. J. Control Autom. Syst. 16 (2) (2018) 630–639.
[21] Y. Hu, B. Liu, Q. Zhou, C. Yang, Recursive extended least squares parameter estimation for wiener nonlinear systems with moving average noises, Circuits Syst. Signal Process. 33 (2) (2014) 655–664.
[22] C. Chen, R. Xiong, W. Shen, A lithium-ion battery-in-the-loop approach to test and validate multiscale dual h infinity filters for state-of-charge and capacity estimation, IEEE Trans. Power Electron. 33 (1) (2018) 332–342.
[23] H. Dong, Z. Wang, S.X. Ding, H. Gao, On h-infinity estimation of randomly occurring faults for a class of nonlinear time-varying systems with fading channels, IEEE Trans. Autom. Control 61 (2) (2016) 479–484.
[24] M. Li, X. Liu, The least squares based iterative algorithms for parameter estimation of a bilinear system with autoregressive noise using the data filtering technique, Signal Process. 147 (2018) 23–34.
[25] X. Zhang, L. Xu, F. Ding, T. Hayat, Combined state and parameter estimation for a bilinear state space system with moving average noise, J. Franklin Inst. 355 (6) (2018) 3079–3103.
[26] X. Wang, F. Ding, Recursive parameter and state estimation for an input nonlinear state space system using the hierarchical identification principle, Signal Process. 117 (2015) 208–218.
[27] M.S. Asif, A. Charles, J. Romberg, C. Rozell, Estimation and dynamic updating of time-varying signals with sparse variations, in: Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2011, pp. 3908–3911.
[28] M.S. Asif, J. Romberg, Sparse recovery of streaming signals using ℓ_1-homotopy, IEEE Trans. Signal Process. 62 (16) (2014) 4209–4223.
[29] D. Angelosante, S.I. Roumeliotis, G.B. Giannakis, Lasso–Kalman smoother for tracking sparse signals, in: Proc. 43rd Asilomar Conference on Signals, Systems and Computers, 2009, pp. 181–185.
[30] A. Charles, M.S. Asif, J. Romberg, C. Rozell, Sparsity penalties in dynamical system estimation, in: Proc. 45th Annual Conference on Information Sciences and Systems, 2011, pp. 1–6.
[31] A.S. Charles, C.J. Rozell, Dynamic filtering of sparse signals using reweighted ℓ_1, in: Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, pp. 6451–6455.
[32] E. Karseras, K. Leung, W. Dai, Tracking dynamic sparse signals with Kalman filters: Framework and improved inference, in: Proc. 10th Int. Conf. on Sampling Theory and Applications (SampTA), 2013, pp. 224–227.
[33] H. Liu, Y. Li, Y. Zhou, T.K. Truong, Sparse Kalman filter, in: Proc. IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP), 2015, pp. 1022–1026.
[34] A.S. Charles, A. Balavoine, C.J. Rozell, Dynamic filtering of time-varying sparse signals via ℓ_1 minimization, IEEE Trans. Signal Process. 64 (21) (2016) 5644–5656.
[35] N. Vaswani, Kalman filtered compressed sensing, in: Proc. 15th IEEE International Conference on Image Processing, 2008, pp. 893–896.
[36] M. Ho, J. Xin, Sparse Kalman filtering approaches to covariance estimation from high frequency data in the presence of jumps, 2016. arXiv: 1602.02185.
[37] M. Partovibakhsh, G. Liu, An adaptive unscented Kalman filtering approach for online estimation of model parameters and state-of-charge of lithium-ion batteries for autonomous mobile robots, IEEE Trans. Control Syst. Technol. 23 (1) (2015) 357–363.
[38] V. Filipovic, N. Nedic, V. Stojanovic, Robust identification of pneumatic servo actuators in the real situations, Forsch. Ingenieurwes. 75 (4) (2011) 183–196.
[39] V. Stojanovic, N. Nedic, Joint state and parameter robust estimation of stochastic nonlinear systems, Int. J. Robust Nonlinear Control 26 (14) (2016) 3058–3074.

[40] J. Hu, Z. Wang, S. Liu, H. Gao, A variance-constrained approach to recursive state estimation for time-varying complex networks with missing measurements, Automatica 64 (2016) 155–162.
[41] C. Nemeth, P. Fearnhead, L. Mihaylova, Sequential monte carlo methods for state and parameter estimation in abruptly changing environments, IEEE Trans. Signal Process. 62 (5) (2014) 1245–1255.
[42] A. Mehmood, M.S. Aslam, N.I. Chaudhary, A. Zameer, M.A.Z. Raja, Parameter estimation for hammerstein control autoregressive systems using differential evolution, Signal Image Video Process. 12 (8) (2018) 1603–1610.
[43] S. Sankaranarayanan, N. Sivakumaran, T. Radhakrishnan, G. Swaminathan, Metaheuristic-based approach for state and process parameter prediction using hybrid grey wolf optimization, Asia-Pac. J. Chem. Eng. 13 (4) (2018) e2215.
[44] A. Srivastava, D.K. Das, A. Rai, R. Raj, Parameter estimation of a permanent magnet synchronous motor using whale optimization algorithm, in: 2018 Recent Advances on Engineering, Technology and Computational Sciences (RAETCS), 2018, pp. 1–6.
[45] Q. Liao, S. Zhou, H. Shi, W. Shi, Parameter estimation of nonlinear systems by dynamic cuckoo search, Neural Comput. 29 (4) (2017) 1103–1123.
[46] N. Nedic, V. Stojanovic, V. Djordjevic, Optimal control of hydraulically driven parallel robot platform based on firefly algorithm, Nonlinear Dyn. 82 (3) (2015) 1457–1473.
[47] V. Stojanovic, N. Nedic, D. Prsic, L. Dubonjic, V. Djordjevic, Application of cuckoo search algorithm to constrained control problem of a parallel robot platform, Int. J. Adv. Manuf.Technol. 87 (9) (2016) 2497–2507.
[48] V. Stojanovic, N. Nedic, A nature inspired parameter tuning approach to cascade control for hydraulically driven parallel robot platform, J. Optim. Theory Appl. 168 (1) (2016) 332–347.
[49] O. Rahbari, N. Omar, Y. Firouz, M.A. Rosen, S. Goutam, P.V.D. Bossche, J.V. Mierlo, A novel state of charge and capacity estimation technique for electric vehicles connected to a smart grid based on inverse theory and a metaheuristic algorithm, Energy 155 (2018) 1047–1058.
[50] L. Ambrogioni, U. Güçlü, E. Maris, M. van Gerven, Estimating nonlinear dynamics with the ConvNet smoother, 2017. arXiv:1702.05243.
[51] N.K. Logothetis, What we can do and what we cannot do with fMRI, Nature 453 (7197) (2008) 869–878.
[52] R.K. Mehra, On the identification of variances and adaptive Kalman filtering, IEEE Trans. Autom. Control 15 (2) (1970). 175–84
[53] R.K. Mehra, Approaches to adaptive filtering, IEEE Trans. Autom. Control 17 (5) (1972). 693–98
[54] B.J. Odelson, M.R. Rajamani, J.B. Rawlings, A new autocovariance least-squares method for estimating noise covariances, Automatica 42 (2) (2006). 303–8
[55] B. Feng, M. Fu, H. Ma, Y. Xia, B. Wang, Kalman filter with recursive covariance estimation–sequentially estimating process noise covariance, IEEE Trans. Ind. Electron. 61 (11) (2014) 6253–6263.
[56] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM J. Imaging Sci. 2 (1) (2009) 183–202.
[57] A. Taylor, J. Hendrickx, F. Glineur, Exact worst-case performance of first-order methods for composite convex optimization, SIAM J. Optim. 27 (3) (2017) 1283–1313.
[58] D. Kim, J. Fessler, Another look at the fast iterative shrinkage/thresholding algorithm (FISTA), SIAM J. Optim. 28 (1) (2018) 223–250.
[59] R. Glowinski, A. Marrocco, Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires, Modélisation Mathématique et Analyse Numérique 9 (R2) (1975) 41–76.
[60] D. Gabay, B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite-element approximations, Comput. Math. Appl. 2 (1) (1976) 17–40.
[61] J. Eckstein, D.P. Bertsekas, On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators, Math. Program. 55 (1–3) (1992) 293–318.
[62] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Found. Trends Mach. Learn. 3 (1) (2011) 1–122.
[63] D.S. Weller, S. Ramani, J.A. Fessler, Augmented Lagrangian with variable splitting for faster non-Cartesian L1-SPIRiT MR image reconstruction, IEEE Trans. Med. Imaging 33 (2) (2014) 351–361.
[64] E. Ghadimi, A. Teixeira, I. Shames, M. Johansson, Optimal parameter selection for the alternating direction method of multipliers (ADMM): quadratic problems, IEEE Trans. Autom. Control 60 (3) (2015). 644–58
[65] S. Ramani, T. Blu, M. Unser, Monte-Carlo SURE: a black-box optimization of regularization parameters for general denoising algorithms, IEEE Trans. Image Process. 17 (9) (2008) 1540–1554.
[66] S. Ramani, D.S. Weller, J.-F. Nielsen, J.A. Fessler, Non-Cartesian MRI reconstruction with automatic regularization via Monte-Carlo SURE, IEEE Trans. Med. Imaging 32 (8) (2013) 1411–1422.
[67] D.S. Weller, S. Ramani, J.-F. Nielsen, J.A. Fessler, Monte Carlo SURE-based parameter selection for parallel magnetic resonance imaging reconstruction, Magn. Reson. Med. 71 (5) (2014) 1760–1770.
[68] H.W. Engl, Discrepancy principles for Tikhonov regularization of ill-posed problems leading to optimal convergence rates, J. Optim. Theory Appl. 52 (2) (1987) 209–215.
[69] X. Hu, T.H. Le, Artifact reduction in EPI with phase-encoded reference scan, Magn. Reson. Med. 36 (1) (1996). 166–71
[70] N. Chen, A.M. Wyrwicz, Removal of EPI Nyquist ghost artifacts with two-dimensional phase correction, Magn. Reson. Med. 51 (6) (2004). 1247–53
[71] S.B. Reeder, E. Atalar, A.Z. Faranesh, E.R. McVeigh, Referenceless interleaved echo-planar imaging, Magn. Reson. Med. 41 (1) (1999) 87–94.
[72] P. Thevenaz, U.E. Ruttimann, M. Unser, A pyramid approach to subpixel registration based on intensity, IEEE Trans. Image Process. 7 (1) (1998) 27–41.
[73] S.Y. Chun, J.A. Fessler, A simple regularizer for B-spline nonrigid image registration that encourages local invertibility, IEEE J. Sel. Top. Signal Process. 3 (1) (2009). 159–69
[74] R.K.-S. Kwan, A.C. Evans, G.B. Pike, MRI simulation-based evaluation of image-processing and classification methods, IEEE Trans. Med. Imaging 18 (11) (1999). 1085–97
[75] K.J. Friston, P. Fletcher, O. Josephs, A. Holmes, M.D. Rugg, R. Turner, Event-related fMRI: characterizing differential responses, NeuroImage 7 (1) (1998) 30–40.
[76] T. Kober, J.P. Marques, R. Gruetter, G. Krueger, Head motion detection using FID navigators, Magn. Reson. Med. 66 (1) (2011). 135–43
[77] G.H. Glover, T.-Q. Li, D. Ress, Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR, Magn. Reson. Med. 44 (1) (2000). 162–7