scientific reports

OPEN

Fast data-driven learning of parallel MRI sampling patterns for large scale problems

Marcelo V. W. Zibetti[1✉], Gabor T. Herman[1,2] & Ravinder R. Regatte[1]

In this study, a fast data-driven optimization approach, named bias-accelerated subset selection (BASS), is proposed for learning efficacious sampling patterns (SPs) with the purpose of reducing scan time in large-dimensional parallel MRI. BASS is applicable when Cartesian fully-sampled k-space measurements of specific anatomy are available for training and the reconstruction method for undersampled measurements is specified; such information is used to define the efficacy of any SP for recovering the values at the non-sampled k-space points. BASS produces a sequence of SPs with the aim of finding one of a specified size with (near) optimal efficacy. BASS was tested with five reconstruction methods for parallel MRI based on low-rankness and sparsity that allow a free choice of the SP. Three datasets were used for testing, two of high-resolution brain images (T_2-weighted images and, respectively, $T_{1\rho}$-weighted images) and another of knee images for quantitative mapping of the cartilage. The proposed approach has low computational cost and fast convergence; in the tested cases it obtained SPs up to 50 times faster than the currently best greedy approach. Reconstruction quality increased by up to 45% over that provided by variable density and Poisson disk SPs, for the same scan time. Optionally, the scan time can be nearly halved without loss of reconstruction quality. Quantitative MRI and prospective accelerated MRI results show improvements. Compared with greedy approaches, BASS rapidly learns effective SPs for various reconstruction methods, using larger SPs and larger datasets; enabling better selection of sampling-reconstruction pairs for specific MRI problems.

Motivation. Magnetic resonance imaging (MRI) is one of the most versatile imaging modalities, it can provide answers to medical questions through the measurements of various properties of the resonant spins in the human body. Unfortunately, the more information we seek from MRI, the longer is the acquisition time[1,2]. This makes the acquisition of high-resolution three-dimensional (3D) volume imaging of the human body time-consuming. Shortening the scan time in MRI is necessary for capturing dynamic processes, quantitative measurements, and for reducing health-care costs and increasing patient comfort. One effective way to reduce scan time is through *undersampling*, in which only part of the total set of measurements, specified by a *sampling pattern* (SP), is acquired. This approach is also called *accelerated MRI*.

The specific content of this paper. We propose and validate a new *data-driven optimization* (DDO) approach to learn the SP in parallel MRI applications. Our focus is on Cartesian 3D high-resolution and 3D quantitative MRI problems in which the data are collected along multiple k-space lines (in the frequency-encoding direction) with the SP a 2D (phase/partition-encoding directions) entity, as illustrated in Fig. 1a. Reconstruction may be be performed as a fully 3D process, but we assume that a Fourier transform is applied in the frequency-encoding direction and the volume is separated into multiple slices for 2D reconstructions. We also tested the method with smaller-size 1D (phase-encoding) SP used in Cartesian 2D acquisitions, as illustrated in Fig. 1b.

The proposed approach is applicable to any parallel MRI method that allows a free selection of the SP, such as compressed sensing (CS)[3–6] and low-rank approaches. Methods that directly recover the k-space elements, such as simultaneous auto-calibrating and k-space estimation (SAKE)[7], low-rank modeling of local k-space neighborhoods (LORAKS)[8], generic iterative re-weighted annihilation filter (GIRAF)[9], and annihilating filter-based

---

[1]Center for Biomedical Imaging, Department of Radiology, New York University School of Medicine, New York, NY 10016, USA. [2]Department of Computer Science, The Graduate Center, City University of New York, New York, NY 10016, USA. ✉email: Marcelo.WustZibetti@nyulangone.org
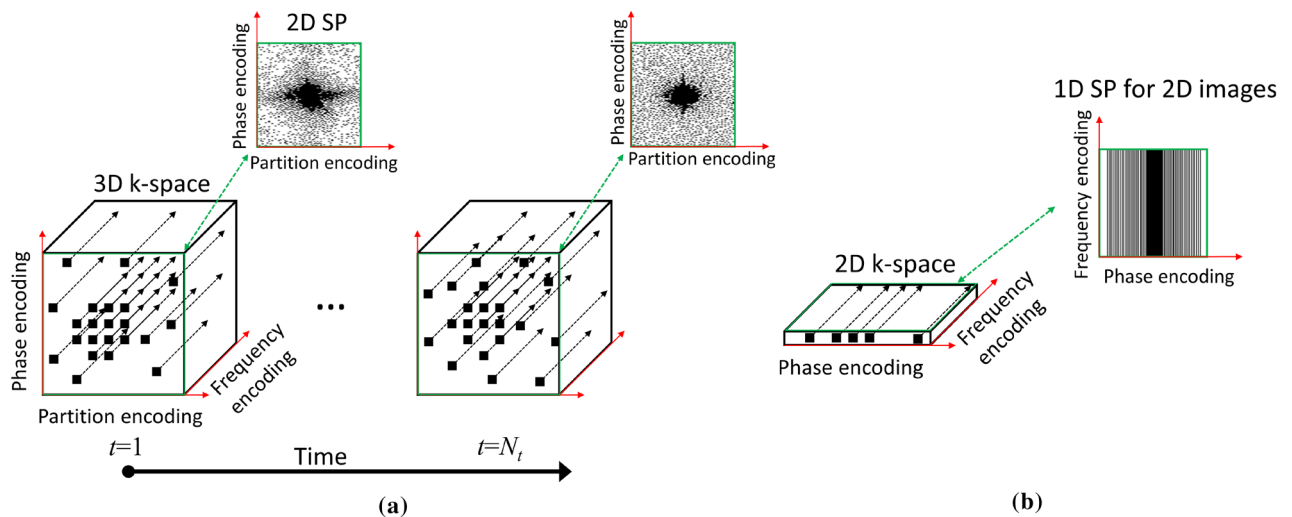

Figure 1. Illustration of the (**a**) 3D + time data collection scheme considered in this work. The sampling pattern is in 2D + time, it comprises the time-varying phase and partition encoding positions, for each of which data are to be collected by the MRI scanner for all frequency encoding positions. Our method can also be applied to (**b**) 2D data collection with fully-sampled lines in the frequency-encoding direction and a 1D sampling pattern denoting phase encoding positions.

low-rank Hankel matrix approach (ALOHA)[10], among others, can be used. We tested the proposed optimization approach for P-LORAKS[11] and three different multi-coil CS approaches with different priors[12]. The contribution of the proposed approach is a new learning algorithm, named bias-accelerated subset selection (BASS), that can optimize large sampling patterns, using large datasets, spending significantly less processing times as compared to previous approaches. Moreover, the SPs optimized by BASS can achieve good image quality with short acquisition times, improving clinical tasks. A very preliminary presentation of our approach is in[13].

Background and purpose. Fast magnetic resonance (MR) pulse sequences for measurements acquisition[1,2,14], parallel imaging (PI) using multichannel receive radio frequency arrays[15–17], and CS[3–6] are examples of advancements towards rapid MRI. PI uses multiple receivers with different spatial coil sensitivities to capture samples in parallel[18], increasing the amount of measurements in the same scan time. Further, undersampling can be used to reduce the overall scan time[15–17]. CS relies on incoherent sampling and sparse reconstruction. With incoherence, the sparse signals spread almost uniformly in the sampling domain, and random-like patterns can be used to undersample the k-space[3–5,19,20].

Successful reconstructions with undersampled measurements, such as PI and CS, use prior knowledge about the true signal to remove the artifacts of undersampling, preserving most of the desired signal. Essentially, the true signal is redundant and can be compactly represented in a certain domain, subspace, or manifold, of much smaller dimensionality[21,22]. Low-rank signal representation[23] and sparse representation[24], are two examples of this kind. Deep learning-based reconstructions have shown that undersampling artifacts can also be separated from true signals by learning the parameters of a neural network from sampled datasets[23,25,26].

The quality of image reconstruction depends on the sampling process. CS is an example of how the SP can be modified[27–29], compared to standard uniform sampling[30], so as to be effective for a specific signal recovery strategy[29,31]. According to pioneering theoretical results[27,32,33], restricted isometry properties (RIP) and incoherence are key for CS. In MRI, however, RIP and incoherence are more like guidelines for designing random sampling[3,5,29] than target properties. New theoretical results[34,35] revisited the effectiveness of CS in MRI; in particular, elucidating that incoherence is not a strict requirement. Also, studies[36,37] show that SPs with optimally incoherent measurements[3] do not achieve the best reconstruction quality, leaving room for effective empirical designs. SPs such as variable density[38–40], Poisson disk[41,42], or even a combination of both[43,44] show good results in MRI reconstruction without relying on optimal incoherence properties.

In many CS-MRI methods, image quality improves when SP is learned utilizing a fully sampled k-space of similar images of particular anatomy as a reference[45–49]. Such adaptive sampling approaches adjust the probability of the k-space points of variable density SP according to the k-space energy of reference images[45–50]. Such SP design methods have been developed for CS reconstructions, but generally they do not consider the reconstruction method to be used.

Statistical methods for optimized design techniques can be used for finding best sampling patterns[51,52]. Experimental design methods, especially using optimization of Cramér-Rao bounds, are general and focus on obtaining improved signal-to-noise ratio (SNR). These approaches were used for fingerprinting[53], PI[54], and sparse reconstructions[51]. They do not consider specific capabilities of the reconstruction algorithm in the design of the SP, even though some general formulation is usually assumed.

In DDO approaches, the SP is optimized for reference images or datasets containing several images of particular anatomy, using a specific method for image reconstruction[55–59]. The main premise is that the optimized

SP should perform well with other images of the same anatomy when the same reconstruction method is used. These approaches can be extended to jointly learning the reconstruction and the sampling pattern, as shown in[60–62]. DDO is applicable to any reconstruction method that accepts various SPs. In[56], DDO for PI and CS-MRI is proposed, the selection of the SP is formulated as a subset selection problem[63,64], which is solved using greedy optimization of an image domain criterion (an extension of[55] for single-coil MRI); see also[57].

Finding an optimal SP via subset selection problem is, in general, an NP-hard problem. Also, each candidate SP needs to be evaluated on a large set of images, which may involve reconstructions with high computational cost. Effective minimization algorithms are fundamental for the applicability of these DDO approaches with large sampling patterns.

Existent subset selection approaches for SP optimization. Commonly used in prior works are the greedy approaches; classified as forward[29,55,65] (increase the number of points sampled in the SP, starting from the empty set), backward[51,65] (reduce the number of points in the SP, from fully sampled), or hybrid[63]. Considering the current SP, greedy approaches test candidates SPs, that are one k-space element different, to be added (or removed). After testing, they add (or remove) the k-space element that provides the best improvement in the cost function[64].

Greedy approaches have a disadvantage regarding computational cost because of the large number of evaluations or reconstructions. Assuming that fully-sampled k-space measurements are of size N, whereas the undersampled measurements are of size $M < N$, and there are N_i images, or data items, used for the learning process, the greedy approach will take $N \times N_i$ reconstructions just to find the best first sampling element of the SP (not considering the next $M - 1$ k-space elements that still have to be computed). This makes greedy approaches computationally unfeasible for large-scale MRI problems. As opposed to this, the approach proposed in this work can obtain a good SP using $50N_i$ to $500N_i$ image reconstructions (for all the M k-space elements of the SP).

The approach in[55] is only feasible because it was applied to one-dimensional (1D) undersampling, such as in Fig. 1b, with a small number of images in the dataset and single-coil reconstructions. The approach was extended to 1D+time dynamic sequences[57] and to parallel imaging[56], but it requires too many evaluations, practically prohibitive for large datasets and large sampling patterns.

A different class of learning algorithms for subset selection[64], not exploited yet by SP learning, use bit-wise mutation, such as Pareto optimization algorithm for subset selection (POSS)[64,66,67]. These learning approaches are less costly per iteration since they evaluate only one candidate SP and accept it if the cost function is improved. POSS is not designed for fast convergence, but for achieving good final results. However, these machine learning approaches can be accelerated if the changes are done smartly and effectively instead of randomly.

Other fast approaches for DDO of SP. Besides the formulation of DDO of SP as a subset selection problem, other approaches have been investigated. The use of deep learning for image reconstruction[23,25,26,68] have been extended to learning the SP. In[60], a probabilistic sampling mask is learned inside the neural network, following by a random generation of SPs. In[62], twice continuously differential models are used to find the SP for variational problems. While these approaches are also faster than[55] to learn the SP, they are less flexible. The parallel MRI methods cited in the Section "The specific content of this paper" cannot be used, and only quadratic cost functions can be optimized. In[61,69,70], parametric formulation of non-Cartesian k-space trajectories are optimized. While being interesting approaches, they cannot be applied to our Cartesian 3D problem described in "The specific content of this paper". Another approach for improving image quality through better sampling is the use of active sampling[71–74], in which the next k-space sampling positions are estimated during the acquisition using the data that have been captured. While promising, this approach requires significant changes within the MRI scanning sequence that are not widely available. As opposed to that, our current approach to find the best (optimized) fixed 3D Cartesian SP for a given anatomy, contrast, and pre-determined reconstruction method, can be included in an accelerated (compressed sensing and parallel) MRI scanning protocol, simply replacing an existent non-optimized SP. For this task, the subset selection formulation of DDO of the SP seems to be the most effective approach for our applications of interest.

Theory
Specification of our aim. Referring to Fig. 1, we use Γ to denote the set of size N comprising in the Cartesian grid all possible (a) time-varying phase and partition encoding positions in the 3D + time data collection scheme or (b) all possible phase encoding positions in the 2D data collection scheme. Our instrument (a multi-coil MRI scanner) can provide measurements related to these sampling positions. Each such "measurement" comprises a fixed number (we denote it by N_s) of measurements values for k-space points, i.e. the points on a line in the frequency-encoding direction for all coils. The measurements for the N positions of Γ are represented as the N_sN-dimensional complex-valued vector $\mathbf{m}$, these are referred to as *fully-sampled measurements*.

Let Ω be any subset (of size $M \leq N$) of Γ; it is referred to as a *sampling pattern* (SP). The undersampled measurements of $\mathbf{m}$, restricted to M positions in Ω, is represented as the N_sM-dimensional complex-valued vector

$$\bar{\mathbf{m}} = \mathbf{S}_\Omega \mathbf{m}, \tag{1}$$

where $\mathbf{S}_\Omega$ is a $N_sM \times N_sN$ matrix is referred to as the *sampling function*. Such $\bar{\mathbf{m}}$ is referred to as the undersampled measurements for the SP Ω. The *acceleration factor* (AF) is defined as N/M. Note that, in practice, the reduction of the scan time depends on the pulse sequence used[2]. We assume here that the acquisition of N_s measurements values for any element of Γ requires the same scan time.

It is assumed that we have a defined *recovery algorithm R* that, for any SP Ω and any undersampled measurements $\bar{\mathbf{m}}$ for that SP, provides an estimate, denoted by $R(\Omega, \bar{\mathbf{m}})$, of the fully-sampled measurements. A method

for finding an efficacious choice Ω in a particular application area is our subject matter. Efficacy may be measured in the following data-driven manner.

Let N_i be the number of images and also the number of fully sampled measurements items (denoted by $\mathbf{m}_1, \ldots, \mathbf{m}_{N_i}$, called the *training measurements*) used in the learning process to obtain an efficacious Ω. Intuitively, we wish to find a SP Ω such that all the measurements $\mathbf{m}_i$, for $1 \le i \le N_i$, are "near" to their respective recovered versions $R(\Omega, \mathbf{S}_\Omega \mathbf{m}_i)$ from the undersampled measurements. Using $f(\mathbf{m}, \mathbf{n})$ to denote the "distance" between two fully-sampled measurements $\mathbf{m}$ and $\mathbf{n}$, we define the *efficacy* of a SP Ω as:

$$F(\Omega) = \frac{1}{N_i} \sum_{i=1}^{N_i} f(\mathbf{m}_i, R(\Omega, \mathbf{S}_\Omega \mathbf{m}_i)). \tag{2}$$

Then the sought-after *optimal sampling pattern* of size M is:

$$\hat{\Omega} = \underset{\substack{\Omega \subset \Gamma \\ s.t.\ |\Omega|=M}}{\arg\min}\, F(\Omega). \tag{3}$$

Models used. Parallel MRI methods that directly reconstruct the images, such as sensitivity encoding method (SENSE)[16,75] and many CS approaches[76], are based on an image-to-k-space forward model, such as

$$\mathbf{m} = \mathbf{FCx} = \mathbf{Ex}, \tag{4}$$

where $\mathbf{x}$ is a vector that represents a 2D+time image of size $N_y \times N_z \times N_t$ (N_y and N_z are horizontal and vertical dimensions, N_t is the number of time frames), $\mathbf{C}$ denotes the coil sensitivities transform mapping $\mathbf{x}$ into multi-coil-weighted images of size $N_y \times N_z \times N_t \times N_c$, with N_c coils. $\mathbf{F}$ represents the spatial Fourier transforms (FT), comprising $N_t \times N_c$ repetitions of the 2D-FT, and $\mathbf{m}$ is the fully sampled measurements, of size $N_y \times N_z \times N_t \times N_c$. The two transforms combine into the encoding matrix $\mathbf{E}$. In 2D+time problems $N = N_y N_z N_t$ and $N_s = N_c$, while in 1D problems $N = N_y$, $N_s = N_z N_c$, and $N_t = 1$. In this work, all vectors, such as $\mathbf{m}$ and $\mathbf{x}$, are represented by bold lowercase letters, and all matrices or linear operators, such as $\mathbf{C}$ or $\mathbf{F}$, are represented by bold uppercase letters.

When accelerated MRI by undersampling is used, the sampling pattern is included in the model as

$$\bar{\mathbf{m}} = \mathbf{S}_\Omega \mathbf{FCx}, \tag{5}$$

where $\mathbf{S}_\Omega$ is the sampling function using SP Ω (same for all coils) and $\bar{\mathbf{m}}$ is the undersampled multi-coil k-space measurements (or k-t-space when $N_t > 1$), with $N_s M$ elements, recalling that the AF is N/M. For reconstructions based on this model, we assumed that a central area of the k-space is fully sampled (such an area is used to compute coil sensitivities with auto-calibration methods, as in[77]).

In parallel MRI methods that recover the multi-coil k-space directly, the undersampling formulation is given by (1) and the image-to-k-space forward model is not used, since one is interested in recovering missing k-space samples using e.g. structured low-rank models[23]. For this, the multi-coil k-space is lifted into a matrix $\mathbf{H} = H(\mathbf{m})$, assumed to be a low-rank structured matrix. Lifting operators $H(\mathbf{m})$ are slightly different across PI methods, exploiting different kinds of low-rank structure[7–11,23].

Once all the samples of the k-space are recovered, the image can be computed by any coil combination[78,79], such as:

$$[\hat{\mathbf{x}}]_n = \sum_{c=1}^{N_c} \mathbf{w}_{n,c}[\mathbf{F}_c^{-1}\mathbf{m}_c]_n, \tag{6}$$

where $\mathbf{m}_c$ is the measurements from coil c, $\mathbf{F}_c^{-1}$ is the inverse 2D-FT for one coil and $\mathbf{w}_{n,c}$ is the weight for spatial position n and coil c.

Reconstruction methods tested. We tested our proposed approach on five different reconstruction methods: Three one-frame parallel MRI methods (SENSE[75], P-LORAKS[11], and PI-CS with anisotropic TV[80,81]) and two multi-frame low-rank and PI-CS methods for quantitative MRI[12].

In P-LORAKS[11,82] the recovery from $\bar{\mathbf{m}}$ produces:

$$R(\Omega, \bar{\mathbf{m}}) = \underset{\substack{\mathbf{m} \\ s.t. \mathbf{S}_\Omega \mathbf{m} = \bar{\mathbf{m}}}}{\arg\min} \left\| H_s(\mathbf{m}) - H_{s,r}(\mathbf{m}) \right\|_F^2, \tag{7}$$

where the operator $H_s(\mathbf{m})$ produces a low-rank matrix and $H_{s,r}(\mathbf{m})$ produces a hard threshold version of the same matrix. P-LORAKS exploits consistency between the sampled k-space measurements and reconstructed measurements; it does not require a regularization parameter. Further, it does not need pre-computed coil sensitivities, nor fully sampled k-space areas for auto-calibration.

SENSE, CS, or low-rank (LR) reconstruction[12] is given by:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\arg\min} \left(\|\bar{\mathbf{m}} - \mathbf{S}_\Omega \mathbf{Ex}\|_2^2 + \lambda P(\mathbf{x}) \right) \approx R_{\mathbf{x}}(\Omega, \bar{\mathbf{m}}), \tag{8}$$

where λ is a regularization parameter. For SENSE, $\lambda = 0$ and no regularization is used. For CS and LR, we looked at the regularizations: $P(\mathbf{x}) = \|\mathbf{Tx}\|_1$, with $\mathbf{T}$ the spatial finite differences (SFD); and low rank (LR), using nuclear-norm of $\mathbf{x}$ reordered as a Casorati matrix $P(\mathbf{x}) = \|\mathbf{M}(\mathbf{x})\|_*$[83].

CS approaches using redundant dictionaries $\mathbf{D}$ in the synthesis models[24,84], given by $\mathbf{x} = \mathbf{Du}$, can be written as:

$$\hat{\mathbf{x}} = \mathbf{D} \cdot \arg\min_{\mathbf{u}} \left(\|\bar{\mathbf{m}} - \mathbf{S}_\Omega \mathbf{EDu}\|_2^2 + \lambda \|\mathbf{u}\|_1 \right) \approx R_{\mathbf{x}}(\Omega, \bar{\mathbf{m}}). \tag{9}$$

A dictionary to model exponential relaxation processes, like T_2 and $T_{1\rho}$, in MR relaxometry problems is the multi-exponential dictionary[12,85]. It generates a multicomponent relaxation decomposition[86]. The approximately-equal symbol $\approx$ is used in (8) and (9), since the iterative algorithm for producing $R_{\mathbf{x}}(\Omega, \bar{\mathbf{m}})$, MFISTA-VA[87] in this paper, may stop before reaching the minimum.

Criteria utilized in this paper. We work primarily with a criterion defined in the multi-coil k-space; see (2) and (3). This criterion is used by parallel MRI methods that recover the k-space components directly in a k-space interpolation fashion (and not in the image-space), such as P-LORAKS[11] and others[7,23,25]. Unless stated otherwise, the $f(\mathbf{m}, \mathbf{n})$ in (2) is

$$f(\mathbf{m}, \mathbf{n}) = \frac{\|\mathbf{m} - \mathbf{n}\|_2^2}{\|\mathbf{m}\|_2^2}. \tag{10}$$

The term $\|\mathbf{m}\|_2^2$ normalizes the error, so that the cost function will not be dominated by datasets with a strong signal.

For image-based reconstruction methods (e.g., SENSE and multi-coil CS) using the model in (4), the $R(\Omega, \mathbf{S}_\Omega \mathbf{m}_i)$ in (2) is replaced by $\mathbf{E}R_{\mathbf{x}}(\Omega, \mathbf{S}_\Omega \mathbf{m}_i)$, as defined, e.g., in (8) and (9). The approach used to obtain the coil sensitivity is part of the method.

Note that (3) can be modified for image-domain criteria as well, such as:

$$\hat{\Omega} = \arg\min_{\substack{\Omega \subset \Gamma \\ s.t. |\Omega| = M}} \left(\frac{1}{N_i} \sum_{i=1}^{N_i} g(\mathbf{x}_i, R_{\mathbf{x}}(\Omega, \mathbf{S}_\Omega \mathbf{m}_i)) \right), \tag{11}$$

where $g(\mathbf{x}, \mathbf{y})$ is a measurement of the distance between images $\mathbf{x}$ and $\mathbf{y}$. In this case, the fully-sampled reference must be computed using a reconstruction algorithm, such as $\mathbf{x}_i = R_{\mathbf{x}}(\Gamma, \mathbf{m}_i)$, and so it is dependent on to the parameters used in that algorithm.

Proposed data-driven optimization. Due to the high computational cost of greedy approaches for large SPs and the relatively low cost of predicting points that are good next candidates, we propose a new learning approach, similar to POSS[64,66,67], but with new heuristics that significantly accelerates the subset selection. For a general description of POSS see[64], Algorithm 14.2.

In our proposed method, similarly to POSS, there is a sequential random selection of the elements to be changed. Differently from POSS, two heuristic rules, named the measure of importance (MI) and the positional constraints (PCs), are used to bias the selection of the elements with the intent to accelerate convergence. This is why our algorithm is named bias-accelerated subset selection (BASS). The MI (defined explicitly in (16) below) is a weight assigned to each element, indicating how much it is likely to contribute to decreasing the cost function. The PCs are rules for avoiding selecting undesirable elements, which may be one of two types: *fixed* or *relative*. Fixed positional constraints rule out the selection of an element because there is some prior reason for *fixing* its value (for example, elements used for auto-calibration are often considered to be such, an area of such elements is illustrated in Fig. 2, top right). Relative positional constraints are inspired by those used in the general combinatorial optimization approach called tabu search (TS)[88], that had been demonstrated to be effective optimization approaches, in which a selection of some elements results in the forbidding of some otherwise legitimate selections in the same iteration. The rules that we have found efficacious in our application are that if an element with high MI is selected, then an adjacent element and also elements that are in complex-conjugated positions should not be selected in the same iteration. However, this does not prevent them to be selected in future iterations.

BASS, aims at finding (an approximation of) the $\hat{\Omega}$ of (3), is described in Algorithm 1. It uses the following user-defined items:

- Ω_{init} is the initial SP for the algorithm. It may be any SP (a Poisson disk, a variable density or even empty SP).
- L is the number of iterations in the training process.
- N is the number of positions in the fully-sampled set Γ.
- M is the desired size of the SP ($M < N$).
- K_{init} is the maximum (initial) number of elements to be added/removed per iteration ($K_{init} < \min(M, N - M)$).
- ρ_r is a function of three positive-integer variables K, M, where ($K < M$), and l, such that $K/M < \rho_r(K, M, l) \leq 1$.
- ρ_a is a function of the positive-integer variables K, M, N, where ($K < N - M$), and l, such that $K/(N - M) < \rho_a(K, M, N, l) \leq 1$.
- **select-remove**$(\Omega, K, \rho_r(K, M, l))$ is a subset of Ω, specified below.
- **select-add**$(\Omega, K, \rho_a(K, M, N, l))$ is a subset of $\Gamma \setminus \Omega$, specified below.
- F is an efficacy function; see (2) with the following.

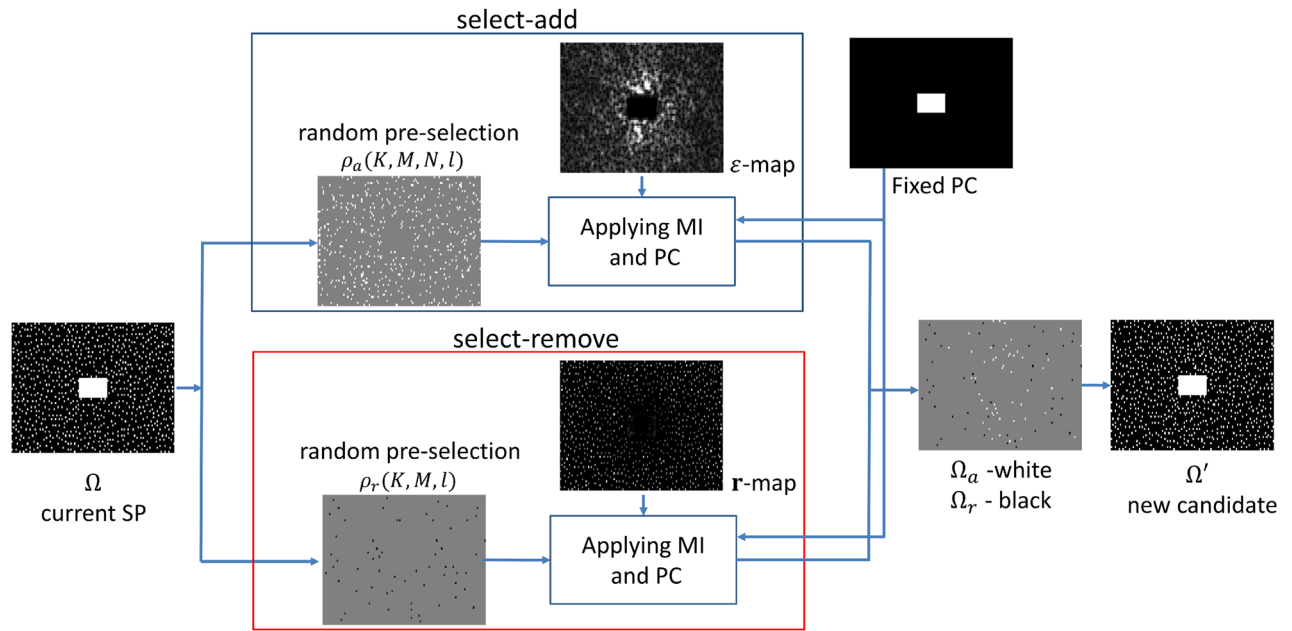 - N_i is the number of items in the training set.

Figure 2. Illustration of the steps used in the functions **select-add** and **select-remove**. First, a random pre-selection is done by Bernoulli trials, using probabilities $\rho_a(K, M, N, l)$ and $\rho_r(K, M, l)$. Later selections are made based on the measurement of importance, using the ε-map and the **r**-map (in which brighter indicates higher values), and the positional constraints (which include identification of the fixed areas). The resulting Ω_a is shown in white and Ω_r in black in the process of composing Ω'. The new candidate Ω' is accepted if the cost function is reduced. These steps are repeated at each iteration.

- $\mathbf{m}_1, \dots, \mathbf{m}_{N_i}$ are the measurements items in the training set.
- R is the recovery algorithm from undersampled measurements.

- α is a reduction factor for the number of elements to be added/removed per iteration ($0 < \alpha < 1$).

Algorithm 1 BASS

1: $\Omega \leftarrow \Omega_{init}$
2: $K \leftarrow K_{init}$
3: $l \leftarrow 1$
4: **while** $l < L$ **do**
5: $\quad \Omega_r \leftarrow$ **select-remove**$(\Omega, K, \rho_r(K, M, l))$
6: $\quad \Omega_a \leftarrow$ **select-add**$(\Omega, K, \rho_a(K, M, N, l))$
7: $\quad \Omega' \leftarrow \Omega_a \cup (\Omega \backslash \Omega_r)$
8: $\quad$ **if** $|\Omega'| \neq M$, **then** $\Omega \leftarrow \Omega'$
9: $\quad$ **if** $|\Omega'| = M$
10: $\quad\quad$ **if** $F(\Omega') \leq F(\Omega)$
11: $\quad\quad$ **then** $\Omega \leftarrow \Omega'$
12: $\quad\quad$ **else** $K \leftarrow \lfloor (K-1)\alpha \rfloor + 1$
13: $\quad l \leftarrow l + 1$
14: **return** Ω

Selection of elements to be added to or removed from the SP. Elements of Ω_a and Ω_r are selected by the functions **select-add** and **select-remove** in similar ways, described in the following paragraphs. First, we point out properties of those selections that ensure the progress of the learning algorithm toward finding an SP of M elements. The properties in question are that if Ω_r, Ω_a, and Ω' are obtained by Steps 5, 6 and 7 of Algorithm 1, then

$$|\Omega_a| = \min\left(\max(M + K - |\Omega|, 0), K\right), \tag{12}$$

$$|\Omega_r| = \min\left(\max(|\Omega| + K - M, 0), K\right), \tag{13}$$

$$\left|\Omega'\right| = |\Omega| + |\Omega_a| - |\Omega_r|. \tag{14}$$

It follows that if $|\Omega| < M$, then $|\Omega_r| < |\Omega_a| = K$ and if $|\Omega| > M$, then $|\Omega_a| < |\Omega_r| = K$. Consequently, if $|\Omega| \neq M$, then

$$\left| \left|\Omega'\right| - M \right| < \left| |\Omega| - M \right|. \tag{15}$$

On the other hand, if $|\Omega| = M$, then $\left|\Omega'\right| = |\Omega|$. Thus, executing Algorithm 1 results in a sequence of $|\Omega|$ that converges to M.

We now define (and illustrate in Fig. 2) **select-add**$(\Omega, K, \rho_a(K, M, N, l))$ and **select-remove**$(\Omega, K, \rho_r(K, M, l))$ in Algorithm 1; they are used in steps 5, 6, and 7. Intuitively, the definitions should be such that the SP Ω' after step 7 is an improved choice as compared to the SP Ω. The number K of elements to be added/removed varies with iteration.

For $1 \leq i \leq N_i$, let $\mathbf{e}_i = \mathbf{m}_i - R(\Omega, \mathbf{S}_\Omega \mathbf{m}_i)$. Here, $\mathbf{e}_i$ is $N_s N$-dimensional vector comprised of N elements, indexed by $1 \leq k \leq N$, each of which is an N_s-dimensional vector, with components indexed by $1 \leq s \leq N_s$; we use $[\mathbf{e}_i]_{k,s}$ to denote the sth component of the kth component of $\mathbf{e}_i$. The kth component of $\mathbf{e}_i$ comprises the N_s measurements related to the kth component of the SP.

For **select-add**, we define a measure of importance (MI) used in this work, for $1 \leq k \leq N$, as

$$\boldsymbol{\varepsilon}_k = \frac{1}{N_i N_s} \sum_{i=1}^{N_i} \frac{\sum_{s=1}^{N_s} \left| [\mathbf{e}_i]_{k,s} \right|^2}{\|\mathbf{m}_i\|_2^2}, \tag{16}$$

referred to as the $\boldsymbol{\varepsilon}$-map. The purpose of **select-add** is to select K elements from $\Gamma \backslash \Omega$ in the following randomly-biased manner. First, an approximately $\rho_a \times (N - M)$ number of elements are randomly pre-selected by Bernoulli trials with ρ_a probability, whose value is the user-provided $\rho_a(K, M, N, l)$ (recall that $K/(N - M) < \rho_a(K, M, N, l) \leq 1$). To have more than K pre-selected points, we need $\rho_a > K/(N - M)$. The K best points of the random pre-selected points will be chosen. The selection starts sequentially with the element with largest MI (the largest ε_k). Once this element is chosen, any other element identified as undesirable by the PC rules is excluded from the randomly pre-selected group, and the selection continues with the element with next largest MI. The chosen K elements are likely to be useful for the aim of (3). The probability ρ_a indirectly controls the bias applied to the selected set. Larger probability implies less randomness and more bias. The probability varies with iteration l.

For **select-remove**, a sequence with number of elements specified in (13), that are in Ω, is generated in the same way, but using $\mathbf{r}_k$ as the MI, instead of ε_k:

$$\mathbf{r}_k = \frac{1}{N_i N_s} \sum_{i=1}^{N_i} \frac{\sum_{s=1}^{N_s} \left| [\mathbf{e}_i]_{k,s} \right|^2 + \delta}{\sum_{s=1}^{N_s} \left| [\mathbf{m}_i]_{k,s} \right|^2 + \delta}, \tag{17}$$

for $1 \leq k \leq N$ with δ a small constant to avoid zero/infinity in the defining of $\mathbf{r}_k$, which is referred to as the $\mathbf{r}$-map. The idea of this MI is that a large reconstruction error in a sampled k-space position k, defined as $\sum_{s=1}^{N_s} \left| [\mathbf{e}_i]_{k,s} \right|^2$, where the expected quadratic value of the element is relatively small, defined as $\sum_{s=1}^{N_s} \left| [\mathbf{m}_i]_{k,s} \right|^2$, renders that element as less important for the SP. The elements of this sequence comprise Ω_r, to be removed from Ω in the process of composing Ω'.

The probability of pre-selecting elements should respect pre-defined ranges, $K/M < \rho_r(K, M, l) \leq 1$ and $K/(N - M) < \rho_a(K, M, N, l) \leq 1$. In order to take advantage of the convergence properties of POSS[64], we argue that the probabilities should be reduced along iterations. In later iterations, according to line 12 of Algorithm 1, we have $K \to 1$ (when $K = 1$, the relative PC are no longer used), then we should also have $\rho_r(K, M, l) \to 1/M$ and $\rho_a(K, M, N, l) \to 1/(N - M)$ for BASS to become like POSS, when the change of elements are purely random. At this point, the same convergence properties of POSS, stated in[64] applies to BASS, given that monotonicity and submodularity are valid. Our recommended choices for lines 6 and 5 of Algorithm 1 are

$$\rho_r(K, M, l) = K/M + ((M - K)/(M \times l)), \tag{18}$$

and

$$\rho_a(K, M, N, l) = K/(N - M) + ((N - M - K)/((N - M) \times l)), \tag{19}$$

with l the iteration index. This results in more bias in the beginning of the iterative process and more randomness in later iterations. Other rules can be used, even constant probabilities ($\rho_r(K, M, l) > 1/M$ and $\rho_a(K, M, N, l) > 1/(N - M)$) for a specific number of iterations. The same PCs were used in **select-add** and **select-remove**.

The expensive part of **select-add** and **select-remove** is the computation of the recoveries given by $R(\Omega, \mathbf{S}_\Omega \mathbf{m}_i)$, but this is done only once per iteration, for all N_i images. These recoveries are also reused to calculate the cost F in line 10 of Algorithm 1. Figure 2 illustrates the steps of these functions using $K = 50$.

Methods

Datasets. In our experiments we utilized three datasets. One dataset, denominated T$_2$-*brain*, contains 40 brain T$_2$-weighted images and k-space measurements from the fast MRI dataset of[89]. Of these, $N_i = 30$ were used for training and $N_v = 10$ for validation. The k-space measurements are of size

$N_y \times N_z \times N_t \times N_c = 320 \times 320 \times 1 \times 16$, and the reconstructed images are $N_y \times N_z \times N_t = 320 \times 320 \times 1$. With this dataset, we tested 2D SPs, of size $N = 320 \times 320$, and 1D SPs, of size $N = 320$ (see Fig. 1). We used 1D SPs with experiments with large numbers of iterations to compare BASS against POSS and greedy approaches. The fast MRI dataset is a public dataset composed of images and k-space data obtained with different acquisitions, not all of them are 3D acquisitions. In this sense, the experiments with 2D SPs in the T$_2$-*brain* dataset are merely illustrative.

The second dataset, T$_{1\rho}$-*brain*, contains T$_{1\rho}$-weighted k-space measurements of the brain, of size $N_y \times N_z \times N_t \times N_c = 128 \times 148 \times 1 \times 20$, and the reconstructed images are $N_y \times N_z \times N_t = 128 \times 148 \times 1$. Unless otherwise stated, $N_i = 65$ were used for training and $N_v = 16$ for validation. This dataset and the next one were all acquired with the Cartesian 3D acquisitions as described in "The specific content of this paper", training and validation sets are composed of data from different individuals.

The third dataset, denominated T$_{1\rho}$-*knee*, contains T$_{1\rho}$-weighted knee images and k-space measurements for quantitative T$_{1\rho}$ mapping, of size $N_y \times N_z \times N_t \times N_c = 128 \times 64 \times 10 \times 15$, and the reconstructed images $N_y \times N_z \times N_t = 128 \times 64 \times 10$ representing the cross-sections of the human knee, and 2D+time SPs of size $N = 128 \times 64 \times 10$. Unless otherwise stated, $N_i = 30$ were used for training and $N_v = 10$ for validation. The k-space measurements for all images are normalized by the largest component. A *reduced-size knee* dataset uses only part of the T$_{1\rho}$-knee dataset. Images are of size $N_y \times N_z \times N_t = 128 \times 64 \times 1$, and $N_i = 5$ and $N_v = 5$. This dataset is used in experiments with a large number of iterations to compare BASS against POSS and greedy approaches for 2D SPs.

Reconstruction methods. For the T$_2$-brain and T$_{1\rho}$-brain datasets, three reconstruction methods were used:

- SENSE[75]: Multi-coil reconstruction, following Eq. (8) with $\lambda = 0$, and minimized with conjugate gradient.
- P-LORAKS[11]: from Eq. (7), with codes available online (https://mr.usc.edu/download/loraks2/).
- CS-SFD[87]: Multi-coil CS with sparsity in the spatial finite differences (SFD) domain, following Eq. (8), and minimized with MFISTA-VA.

SENSE was used only for 1D SP comparisons between BASS, POSS and greedy approaches.

For the T$_{1\rho}$-weighted knee dataset, we used different methods:

- CS-LR[12]: Multi-coil CS using nuclear-norm of the vector **x** reordered as a Casorati matrix $P(\mathbf{x}) = \|\mathbf{M}(\mathbf{x})\|_*$ and minimized with MFISTA-VA.
- CS-DIC[12]: Multi-coil CS using synthesis approach following Eq. (9), using **D** as a multi-exponential dictionary[85], and minimized with MFISTA-VA.

CS-SFD, CS-LR, and CS-DIC need a fully-sampled area for auto-calibration of coil sensitivities using ESPIRiT[77]. P-LORAKS does not use auto-calibration. All experiments were performed in Matlab, codes used in this manuscript are available in https://cai2r.net/resources/software/data-driven-learning-sampling-pattern.

The regularization parameter (the λ in (8) and (9)) required in CS-SFD, CS-LR, and CS-DIC was optimized independently for each type of SP (Poisson disk, variable density, combined variable density and Poisson disk, adaptive SP, or optimized) and each AF, using the same training data. The parameters of the recovery method R are assumed to be fixed during the learning process of the SP.

Optimizing parameters of Poisson disk, variable density, and adaptive SPs. Grid optimization with 50 realizations of each SP was performed, changing the parameters used to generate these SPs, to obtain the best realization of these SPs, which corresponds to the one that minimizes $F(\Omega)$. This approach is the one used in[56] for minimizing $F(\Omega)$. Poisson disk and variable density codes used in the experiments are at https://github.com/mohakpatel/Poisson-Disc-Sampling and http://mrsrl.stanford.edu/~jycheng/software.html. Combined Poisson disk and variable density SP from[44] and adaptive SPs from[45] were also used for comparison. The spectrum template obtained from the same training data was used with adaptive SPs. All these approaches can be considered data-driven approaches because optimization of the parameters to generate the SP was performed. They all have a fixed computational cost of $50N_i$ image reconstructions (nearly the same computational cost as BASS).

Evaluation of the error. The quality of the results obtained with the SP was evaluated using the normalized root mean squared error (NRMSE):

$$\mathrm{NRMSE}\left(\{\mathbf{m}_i\}_{i=1}^{N_v}, \left\{\hat{\mathbf{m}}_i\right\}_{i=1}^{N_v}\right) = \sqrt{\sum_{i=1}^{N_v} \frac{\|\mathbf{m}_i - \hat{\mathbf{m}}_i\|_2^2}{\|\mathbf{m}_i\|_2^2}}. \tag{20}$$

When not specified, the NRMSE shown was obtained from k-space on the validation set; results using image-domain and the training set are also provided, as is structural similarity (SSIM)[90] in some cases.
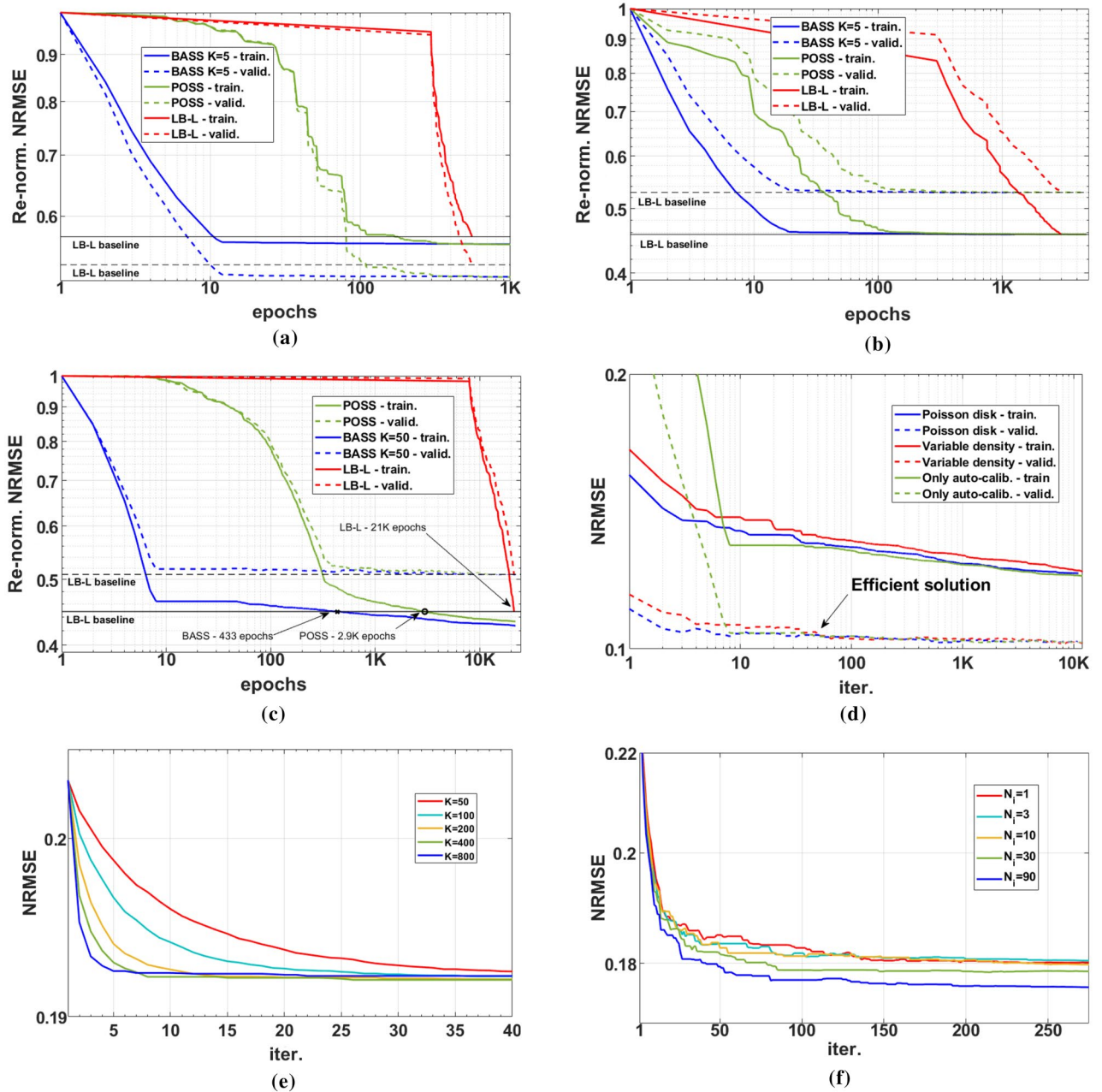
Figure 3. Convergence curves for BASS. Comparison against POSS and the greedy approach LB-L in (**a**) 1D SP using SENSE, (**b**) 1D SP using CS-SFD, and (**c**) 2D SP using CS-SFD. (**d**) Comparing various initial SPs. (**e**) Comparing various K_{init}s. (**f**) Comparing various training sizes.

Results

Illustration of the convergence and choice of parameters. In Fig. 3a–c we compare BASS against POSS[66] and the greedy approach "learning-based lazy" (LB-L)[56], adapted to the cost function in (2). The resulting NRMSEs are re-normalized by the initial values and show the difference in computational cost and quality between the approaches. Plots are scaled logarithmically in epochs (in each "epoch" all the images are reconstructed once). In Fig. 3a, it is shown the performance of the learning methods with 1D SP using T$_2$-*brain* dataset and SENSE reconstruction, with AF = 4. In this example, BASS was faster than POSS and LB-L. Also, BASS and POSS obtained nearly same quality results, superior to LB-L. In Fig. 3b, the performance of the learning methods was tested in the same setup, but using CS-SFD reconstruction, with AF = 4. In this example, BASS was faster than POSS and LB-L, but all methods obtained nearly the same quality results. In Fig. 3c, the methods were compared with CS-SFD with the *reduced-size knee* dataset and 2D SPs, starting with the auto-calibration area and AF = 15. In this example BASS found a solution with same quality in the training set using only 433 epochs, around 50 times faster than LB-L (~ 21,000 epochs). Also, BASS and POSS can go on minimizing the cost function beyond the stopping point of LB-L finding even better SPs.
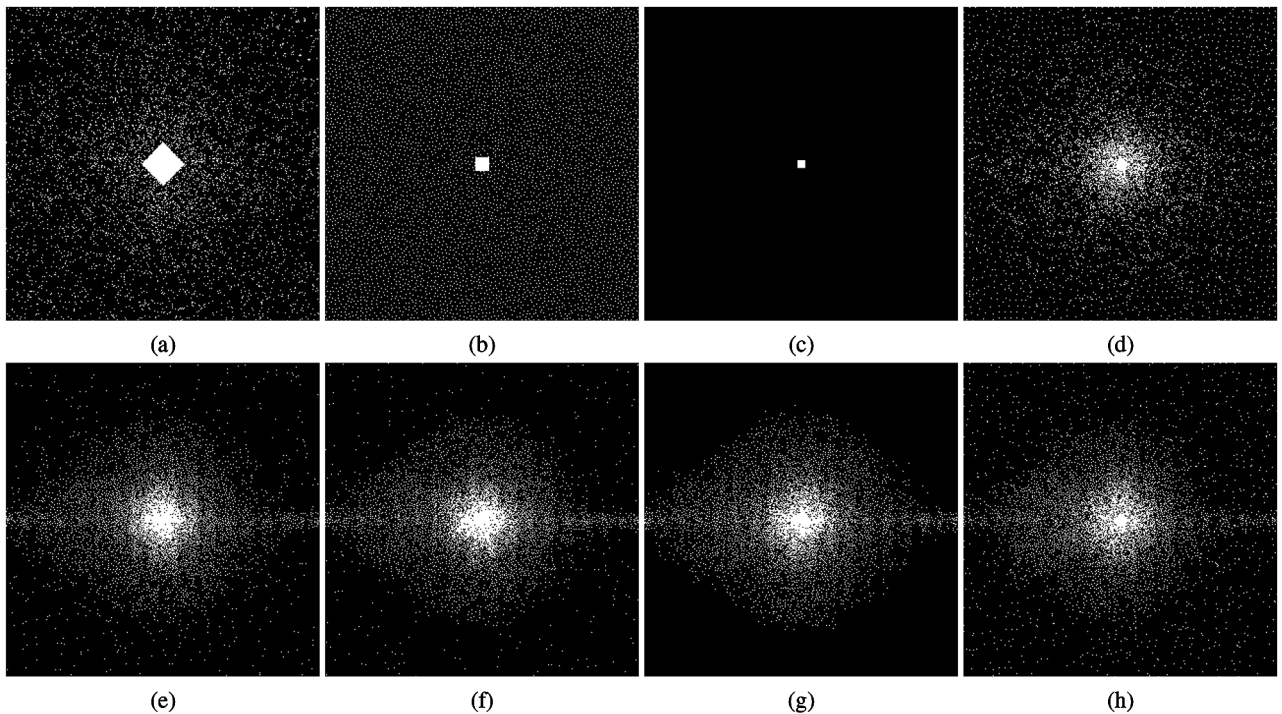
9

Figure 4. Efficient solutions produced for P-LORAKS with AF = 16 and initial SPs (**a**) variable density (VD), (**b**) Poisson disk (PD), (**c**) an SP that is empty except for a small central area (CA), and (**d**) adaptive SP. The corresponding efficient solutions are the SPs in (**e**) for VD (NRMSE = 0.196), in (**f**) for PD (NRMSE = 0.195), in (**g**) for CA (NRMSE = 0.194), and in (**h**) for adaptive SP (NRMSE = 0.199).

Figure 3d, demonstrates the performance of BASS for various initial SPs (same experimental setup as for Fig. 3c, with K_{init}=50). The improvement observable in the validation set ends quickly, at iteration 50 in this example. There is an arrow in the figure pointing to an efficient solution. Such a solution is obtained after relatively few iterations, during which most of the significant improvement observable with validation data has already happened. Iterating beyond this point essentially leads to marginal improvement, observable only with the training data.

In Fig. 3e we see the results of the learning process for the training data according to the parameters K_{init} for CS-LR, AF = 20, using the *knee* dataset, with $N_i = 30$ and $N_v = 10$. Note that large K_{init} performs better than small K_{init} in terms of speed of convergence in the beginning of the learning process. Over time, K reduces from K_{init} towards $K = 1$.

The importance of large and diverse datasets to generate the learned sampling pattern for the specific class of images is illustrated in Fig. 3f, showing the convergence of the learning process with the validation set, in NRMSE. We used training sets of 1, 3, 10, 30, and 90 images. The validation sets were composed of the same 20 images, not used in any of the training sets.

The robustness of an efficient solution in the presence of variable initial SP is illustrated in Fig. 4. Figure 4a–d show three initial SPs: variable density (VD), Poisson disk (PD), empty except for a small central area (CA), and adaptive SP. Using 200 iterations of BASS for P-LORAKS with these initial SPs, corresponding efficient SPs were obtained; shown in Fig. 4e–h. There are minor differences among them (around 1% difference in NRMSE), but the central parts of the SPs are very similar.

Performance with various reconstruction methods. BASS improves NRMSE in image space for fixed AFs when compared with the other SPs for the four reconstruction methods tested with 2D+time SPs. Figure 5a shows the NRMSE obtained by P-LORAKS with T_2-brain dataset, comparing variable density SPs, Poisson disk SPs, adaptive SPs, combined variable density with Poisson disk SPs, and the optimized SPs. Figure 5b shows the NRMSE obtained by CS-SFD with T_2-brain dataset. Figure 5c,d show P-LORAKS and CS-SFD with $T_{1\rho}$-brain, dataset. Figure 5e shows the NRMSE obtained by CS-LR with $T_{1\rho}$-knee dataset. Figure 5f shows the NRMSE obtained by CS-DIC with $T_{1\rho}$-knee dataset. All SP had their parameters optimized for each reconstruction method, dataset, and AF.

Figure 6 illustrates on the T_2-brain dataset how the optimized SPs improve the reconstructed images with P-LORAKS (for AF = 9) and CS-SFD (for AF = 16) against combined variable density and Poisson disk (VD + PD). The P-LORAKS methods had a visible improvement in SNR, the CS-SFD methods became less smooth with some structures more detailed. However, some structured error can still be seen in the error maps. Figure 6 also illustrates that optimized SPs are different for the two reconstruction methods, even when using the same images for training. Figure 7 illustrates on the images obtained with the $T_{1\rho}$-brain dataset with P-LORAKS (for AF = 5) and CS-SFD (for AF = 6), comparing optimized SP with variable density and adaptive SP.
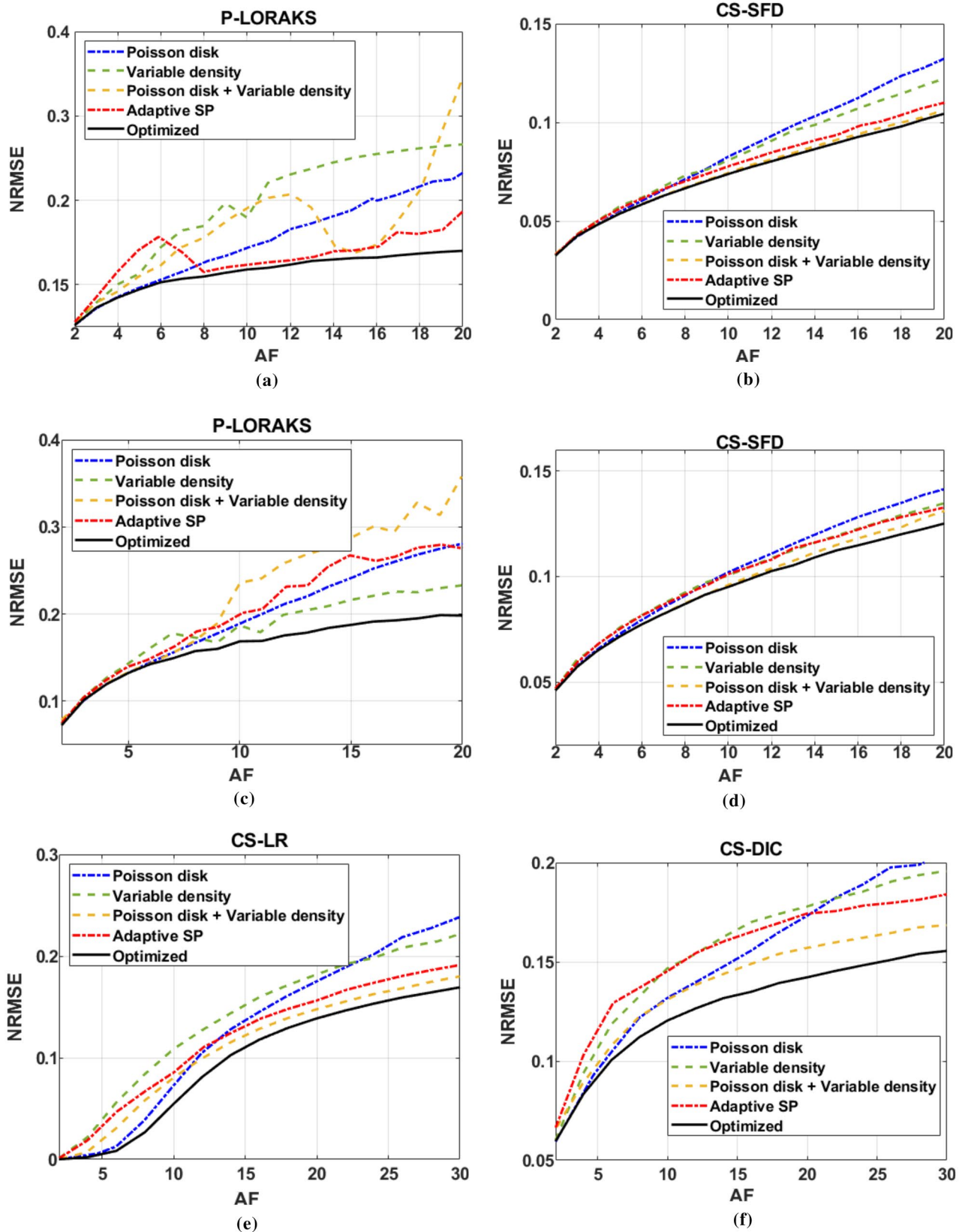
Figure 5. NRMSE (lower is better): (**a**) for P-LORAKS and (**b**) CS-SFD for T_2-*brain* dataset, (**c**) for P-LORAKS and (**d**) CS-SFD for $T_{1\rho}$-*brain*, dataset, and (**e**) for CS-LR, and (**f**) CS-DIC for $T_{1\rho}$-*knee* dataset. Variable density, Poisson disk, adaptive SP, and combined variable density with Poisson disk are compared with optimized SP (obtained by BASS) for various AFs.
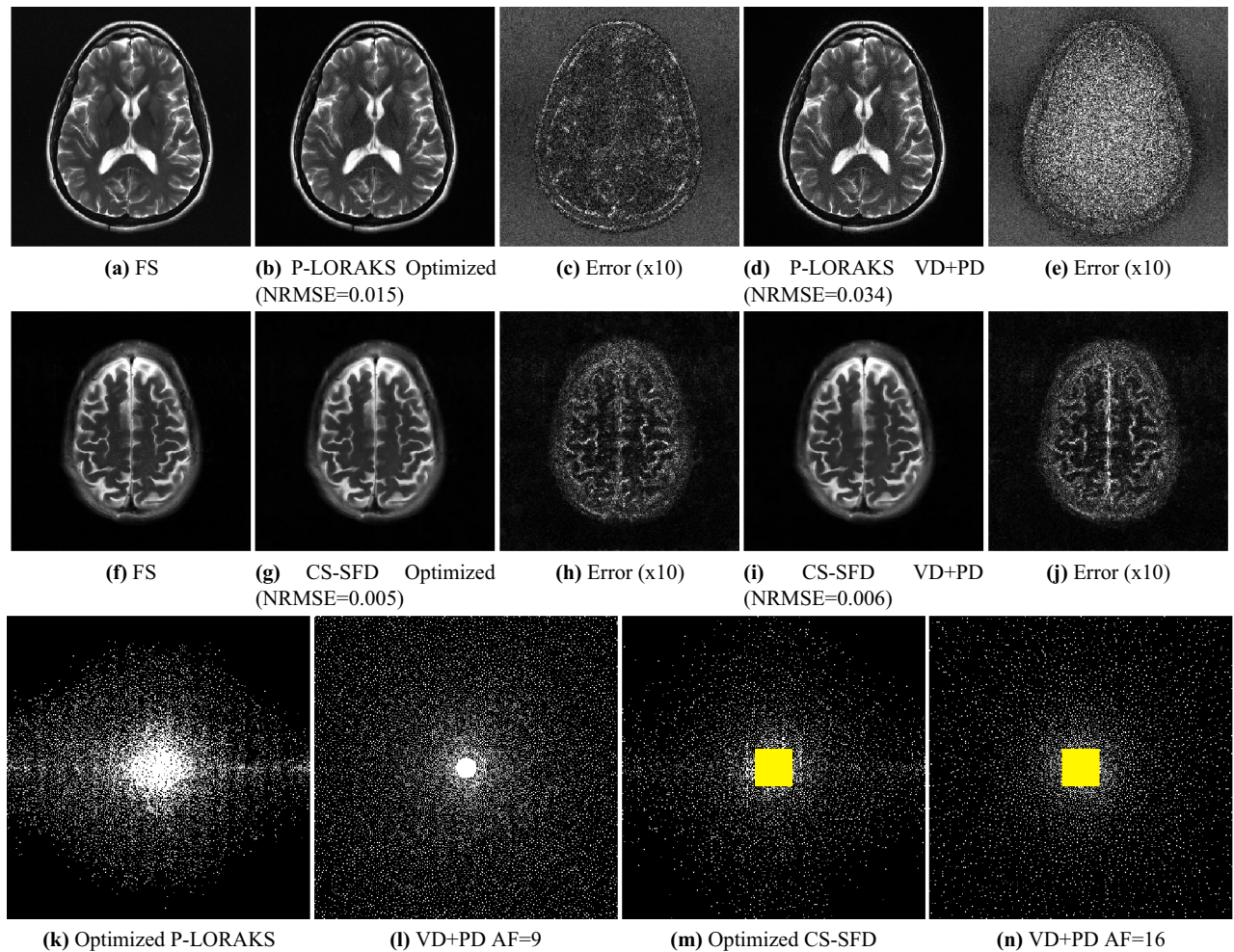
(a) FS   **(b)** P-LORAKS Optimized (NRMSE=0.015)   **(c)** Error (x10)   **(d)** P-LORAKS VD+PD (NRMSE=0.034)   **(e)** Error (x10)

(f) FS   **(g)** CS-SFD Optimized (NRMSE=0.005)   **(h)** Error (x10)   **(i)** CS-SFD VD+PD (NRMSE=0.006)   **(j)** Error (x10)

(k) Optimized P-LORAKS   **(l)** VD+PD AF=9   **(m)** Optimized CS-SFD   **(n)** VD+PD AF=16

Figure 6. Visual results of the T_2-brain dataset are shown in this figure. For P-LORAKS, fully sampled (FS) reference is shown in (**a**), and images using optimized SP and combined variable density and Poisson disk (VD+PD) SP with AF = 9 are shown in (**b**) and (**d**). Error maps between FS and P-LORAKS are shown in (**c**) for Optimized SP and (**e**) for VD + PD SP. For CS-SFD, FS reference is shown in (**f**), and images using optimized SP and VD + PD SP with AF = 16 are shown in (**g**) and (**i**). Error maps between FS and CS-SFD are shown in (**h**) for Optimized SP and (**j**) for VD+PD SP. The optimized and VD+PD SPs with AF = 9 for P-LORAKS are shown in (**k**) and (**l**), while the ones with AF = 16 used for CS-SFD, with the highlighted central-square auto-calibration region, are shown in (**m**) and (**n**).

In Fig. 8, visual results with the $T_{1\rho}$-knee dataset illustrate the improvement due to using an optimized SP as compared to using combined variable density and Poisson disk SP, for both CS-LR and CS-DIC. We also see that the optimized SPs are different for the two reconstruction methods. Note that both optimized k-t-space SPs have a different sampling density over time (first, middle, and last time frames are shown), being more densely sampled at the beginning of the relaxation process. The auto-calibration region is in the first frame.

BASS with a different criterion. We illustrate that our proposed optimization approach is also efficacious with different criteria. In some applications, one may desire the best possible image quality, regardless of k-space measurements. Here we discuss the use of BASS to optimize the SSIM of[90], an image-domain criterion. For that, the task in (3) of finding the minimizer of $F(\Omega)$ in (2), used in line 10 of the Algorithm 1, is replaced by finding the minimizer in (11), with $g(\mathbf{x}, \mathbf{y})$ the negative of the SSIM. In Fig. 9 we compare the optimization of SSIM with that of NRMSE, using P-LORAKS on the T_2-*brain* dataset, AF = 10, starting with the Poisson disk SP.

$T_{1\rho}$ mapping. We illustrate the performance of the optimized SPs for $T_{1\rho}$ mapping. We compare the optimized SP against Poisson disk SP, previously used in[12], for CS-LR. The SP and reconstructed images correspond to the cross-section of the knee, of size $N_y \times N_z \times N_t = 128 \times 64 \times 10$, the $T_{1\rho}$ mapping is performed in the cartilage region on the longitudinal plane (in-plane) of the recovered 3D volume. The 3D+time volume has $N_x \times N_y \times N_z \times N_t = 256 \times 128 \times 64 \times 10$ voxels, where $N_x = 256$ corresponds to the samples in the frequency-encoding direction, field-of-view of 130 mm × 130 mm × 130 mm, and in-plane (rounded) resolution of 0.5 mm × 1 mm, slice thickness of 2 mm, and 10 frames. In Fig. 10 we illustrate the results with $T_{1\rho}$ mapping in the knee, particularly around the cartilage region. In Fig. 10a–c we show some illustrative $T_{1\rho}$ maps.
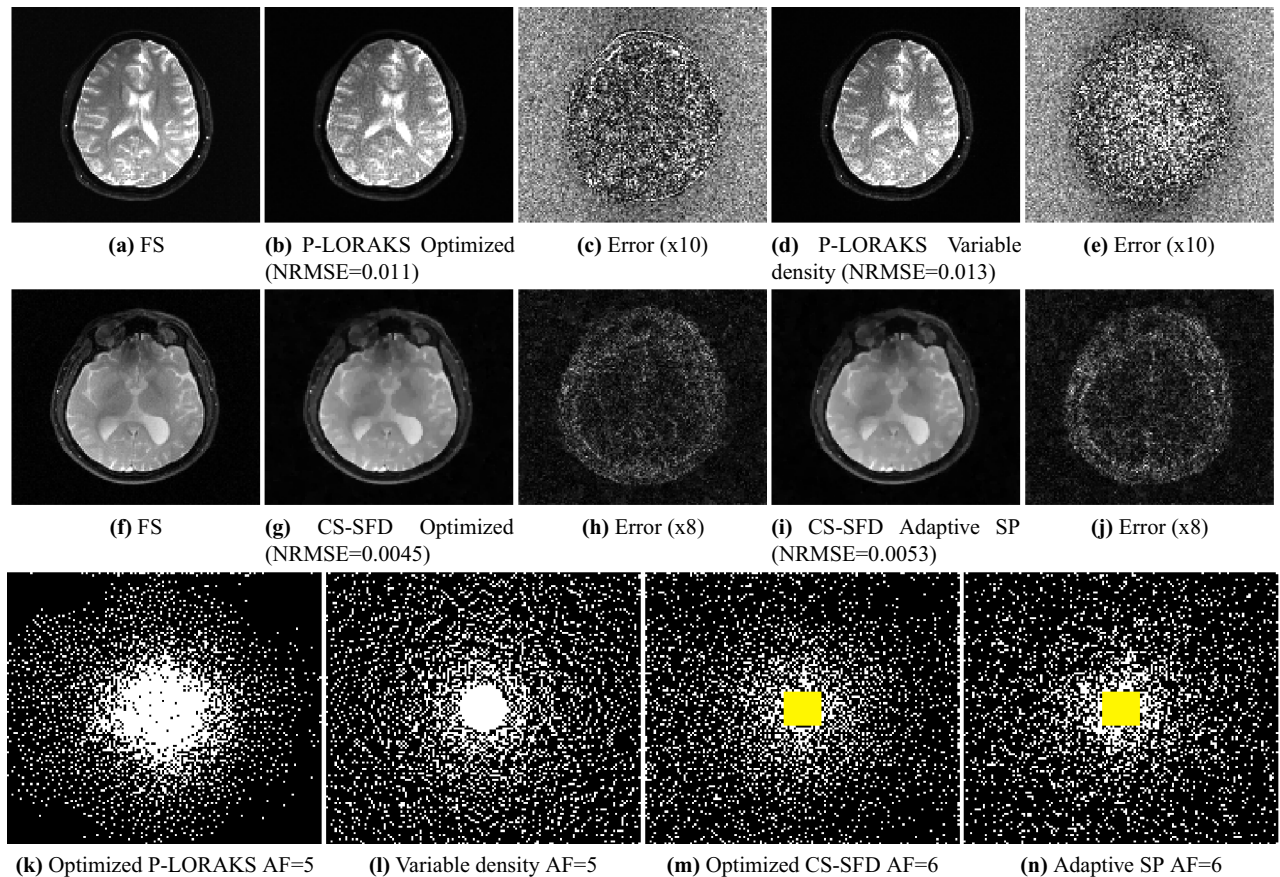
Figure 7. Visual results of the $T_{1\rho}$-brain dataset are shown in this figure. For P-LORAKS, fully sampled (FS) reference is shown in (**a**), and images using optimized SP and variable density SP with AF = 5 are shown in (**b**) and (**d**). Error maps between FS and P-LORAKS are shown in (**c**) for Optimized SP and (**e**) for variable density SP. For CS-SFD, FS reference is shown in (**f**), and images using optimized SP and adaptive SP with AF = 6 are shown in (**g**) and (**i**). Error maps between FS and CS-SFD are shown in (**h**) for Optimized SP and (**j**) for adaptive SP. The optimized and variable density SP with AF = 5 for P-LORAKS are shown in (**k**) and (**l**), while the optimized and adaptive SPs with AF = 6 used for CS-SFD, with the highlighted central-square auto-calibration region, are shown in (**m**) and (**n**).

In Fig. 10d we show the NRMSE for different acceleration factors, considering 10 slices containing the knee cartilage. In Fig. 10e,f, we show the point-wise errors of the $T_{1\rho}$ maps.

Prospective accelerated scans. We tested the optimized SP obtained with BASS in prospective CS scans, in Fig. 11. For an explanation of the usage of the word "prospective" in MRI, see[6]. We used the knee dataset for training the SP for CS-SFD at AF = 4. The images of size $N_y \times N_z \times N_t = 128 \times 64 \times 1$ used for training compose the cross-session of the 3D volumes. Displayed images correspond to the longitudinal view of one slice of a 3D volume (which has size $N_x \times N_y \times N_z = 256 \times 128 \times 64$), with in-plane resolution of 0.5 mm × 1 mm and slice thickness of 2 mm. The 15-channel coil measurements was obtained with the $T_{1\rho}$ pulse sequence used in[12], which is a $T_{1\rho}$ magnetization prepared fast gradient-echo sequence[2].

Discussion

The proposed approach delivers efficacious sampling patterns for high-resolution or quantitative parallel MRI problems. Compared to previous greedy approaches for parallel MRI, as in[56,57], BASS is able to optimize much larger SPs, using larger datasets, spending less computational time than greedy approaches (Fig. 3a–c. Greedy approaches test considerably more candidates SPs before updating the SP. They are computationally affordable only for 1D undersampling or small 2D SPs, but they were inferior to BASS in computational time and imaging quality. Note that computational time for each epoch (or iteration for BASS) depends on the time of the implemented reconstruction algorithm. For CS-DIC, the reconstruction of each slice of the knee dataset takes 16.1 s on an NVIDIA Tesla V100 GPU, while for CS-LR it takes 9.6 s. For the T_2-brain dataset, P-LORAKS takes 110.2 s on a CPU Intel Skylake 2.4 GHz, SENSE takes 0.5 s and CS-SFD takes 6.3 s on an NVIDIA Tesla V100 GPU. We estimate approximate 1.5 computational years running LB-L (estimated 263 K epochs for learning) for 2D SP with $N = 320 \times 320$ using CS-SFD on the T_2-brain dataset (considering only 30 training images) to obtain nearly the same SP that BASS finds in 33 hours (if 500 epochs or iterations are used) in an NVIDIA GPU.
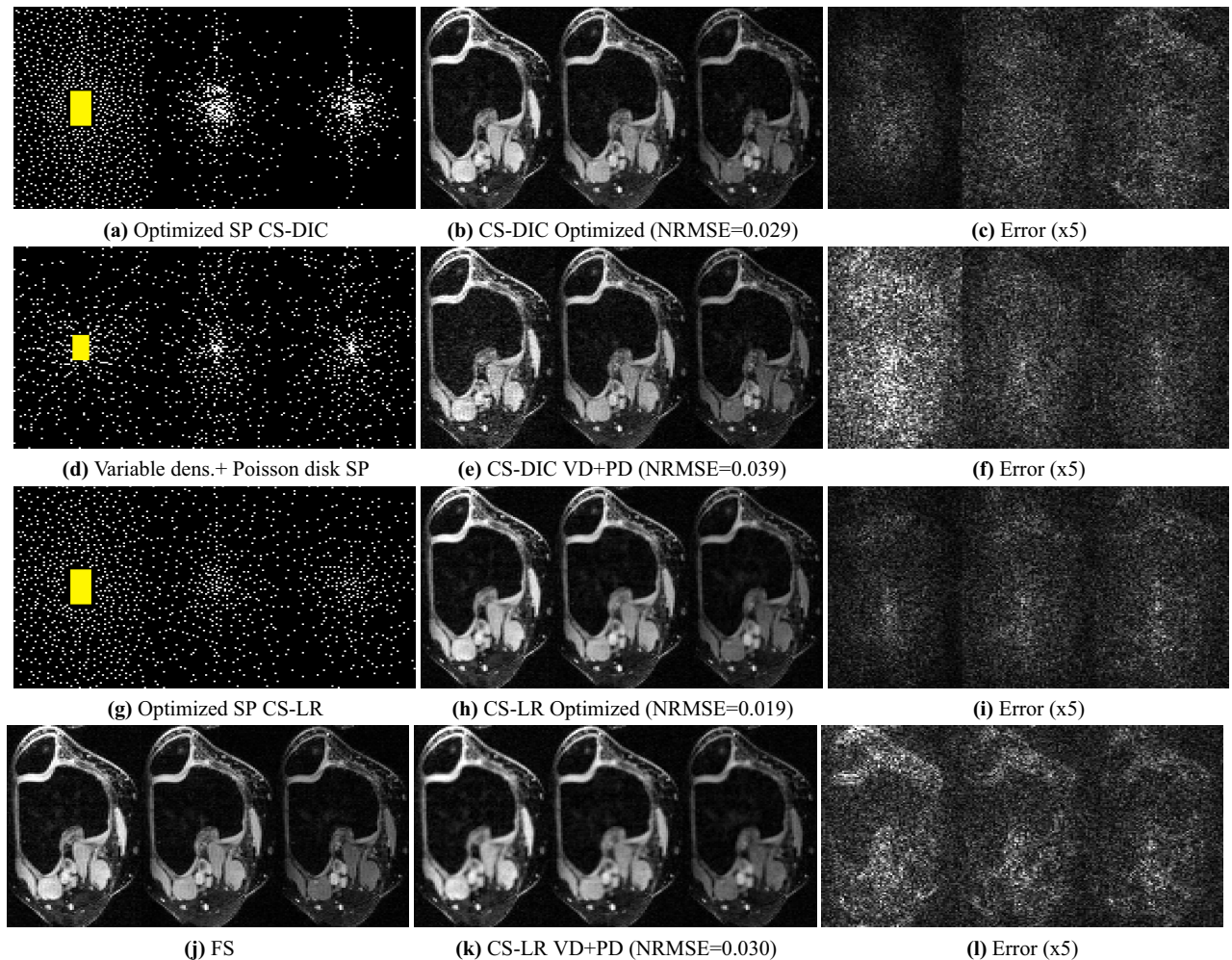
(a) Optimized SP CS-DIC   **(b)** CS-DIC Optimized (NRMSE=0.029)   **(c)** Error (x5)

(d) Variable dens.+ Poisson disk SP   **(e)** CS-DIC VD+PD (NRMSE=0.039)   **(f)** Error (x5)

(g) Optimized SP CS-LR   **(h)** CS-LR Optimized (NRMSE=0.019)   **(i)** Error (x5)

(j) FS   **(k)** CS-LR VD+PD (NRMSE=0.030)   **(l)** Error (x5)

Figure 8. Three frames for different relaxation times of the knee dataset, when AF $=$ 18 was used, reconstructed with CS-DIC (**b**) and (**e**) and with CS-LR (**h**) and (**k**): compare these with the corresponding fully sampled (FS) measurements in (**j**), where the corresponding magnitude of the errors are in (**c,f,i,l**). Combined variable density and Poisson disk SP (**d**) and BASS optimized SPs for CS-DIC (**a**) and CS-LR (**g**) are also shown. Central auto-calibration area is highlighted in yellow.

The proposed approach for subset selection is effective because it uses a smart selection of new elements in the SP updating process. Candidates that are most likely to reduce the cost function are tried first. The obtained efficient solution may have minor differences depending on the initial SP (Fig. 4), but the optimized SPs tend to have the same final quality if more iterations are used (Fig. 3d). Adding and removing multiple elements at each iteration is beneficial for fast convergence at the initial iterations (Fig. 3e).

The cost function in (2) evaluates the error in k-space, not in the image domain. This may not be sufficiently flexible because it does not allow the specification of regions of interest in the image domain. Nevertheless, improvements measured by the image-domain criteria NRMSE were observed (Fig. 5). In different MRI applications other criteria than (2) may be desired. The proposed algorithm can be used for other criteria, such as the SSIM (Fig. 9).

The optimized SP varies with the reconstruction method (Figs. 6, 7 and 8) or with the optimization criterion (Fig. 9): thus sampling and reconstruction should be matched. This concept of matched sampling-reconstruction indicates that comparing different reconstruction methods with the same SP is not a fair approach, instead each MRI reconstruction method should be compared using its best possible SP. Note that the optimized SP improved the NRMSE by up to 45% in some cases (Fig. 5).

The experiments also show that optimizing the SP is more important at higher AFs. As seen in Fig. 5, the optimization of the SP flattened the curves of the error over AF, achieving a lower error with the same AF. For example, P-LORAKS with optimized SP at AF = 20 obtained the same level of NRMSE as with variable density SP at AF = 6, while CS-LR with optimized SP at AF = 30 obtained the same level as with Poisson disk SP at AF = 16, even after optimizing the parameters used to generate the Poisson disk SP. Thus it is possible to double the AF by optimizing the SP. Variable sampling rate over time is advantageous for $T_{1\rho}$ mapping as seen in[91]; it is interesting that the algorithm learned this, as shown in Figs. 8 and 10. It is also important to clarify that the results shown for variable density, Poisson disk, combined variable density and Poisson disk, and adaptive SP are the best obtained among a parameter optimization process spending 50 epochs. If a simple guess of the parameters for these SPs
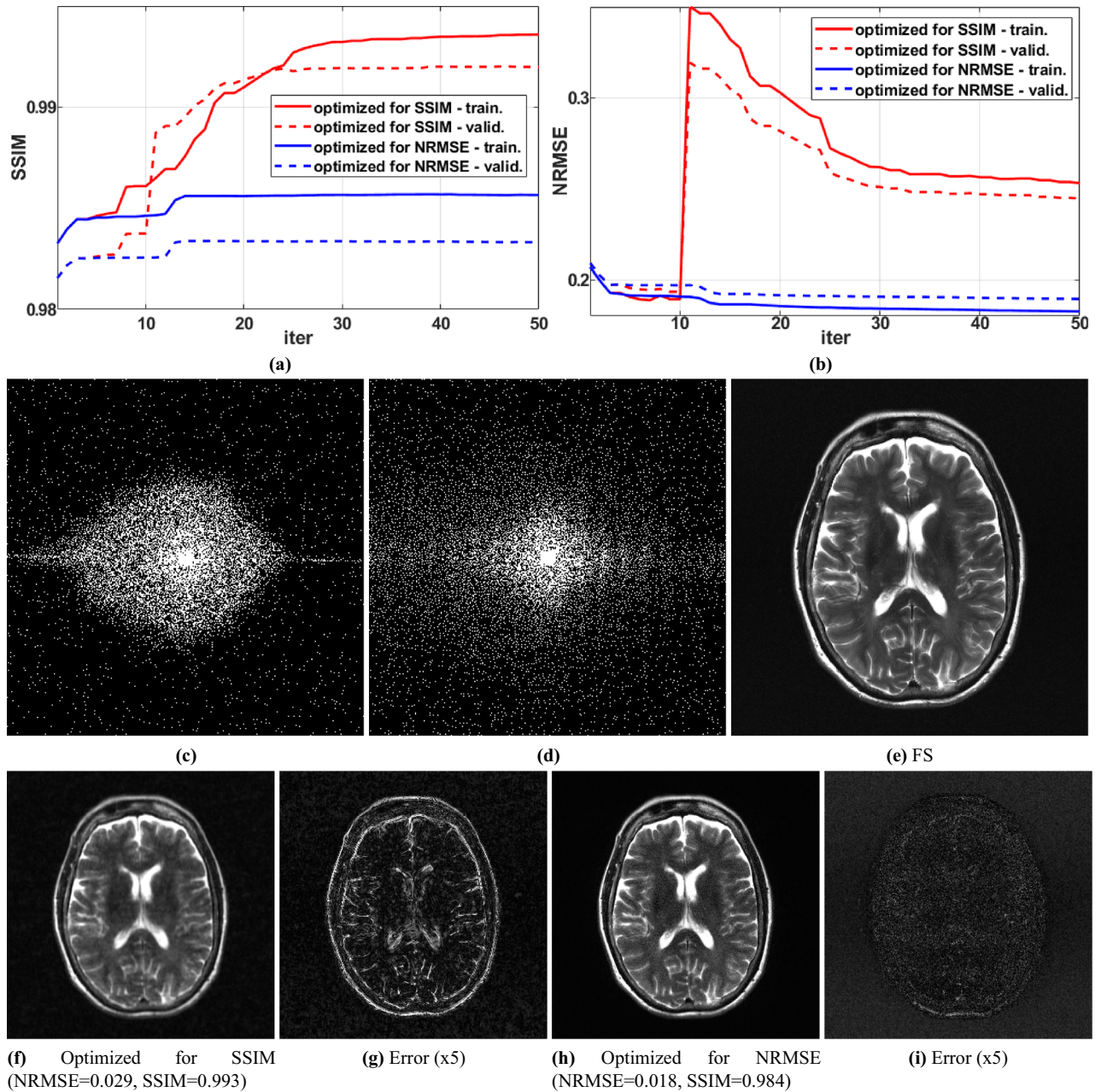
Figure 9. Comparing BASS in optimizing SSIM (higher is better) and NRMSE (lower is better). (**a**) SSIM and (**b**) NRMSE along the iterations, (**c**) SP obtained by optimizing SSIM, (**d**) SP obtained by optimizing NRMSE, and some visual results with fully sampled reconstruction in (**e**), example of images with SP obtained by optimizing SSIM in (**f**) and NRMSE in (**h**), and their error maps (**g**) and (**i**).

is used, then the performance of these SPs can be poor. In contrast, BASS found efficient SPs spending the same computational cost or less than that (10~50 epochs in Fig. 3a–d).

The lower computational cost and rapid convergence speed of BASS bring the advantage of learning the optimal SP for various reconstruction methods considering the same anatomy. Thus one can have a better decision on which matched sampling and reconstruction is the most effective for specific anatomy and contrast at the desired AF. Many questions regarding the best way to sample in accelerated MRI can be answered with the help of machine learning algorithms such as BASS. Learned SPs are key elements in making higher AFs available in clinical scanners for translational research.

Conclusion

We proposed a data-driven approach for learning the sampling pattern in parallel MRI. It has a low computational cost and converges quickly, enabling the use of large datasets to optimize large sampling patterns, which is important for high-resolution Cartesian 3D-MRI and quantitative and dynamic MRI applications. The approach
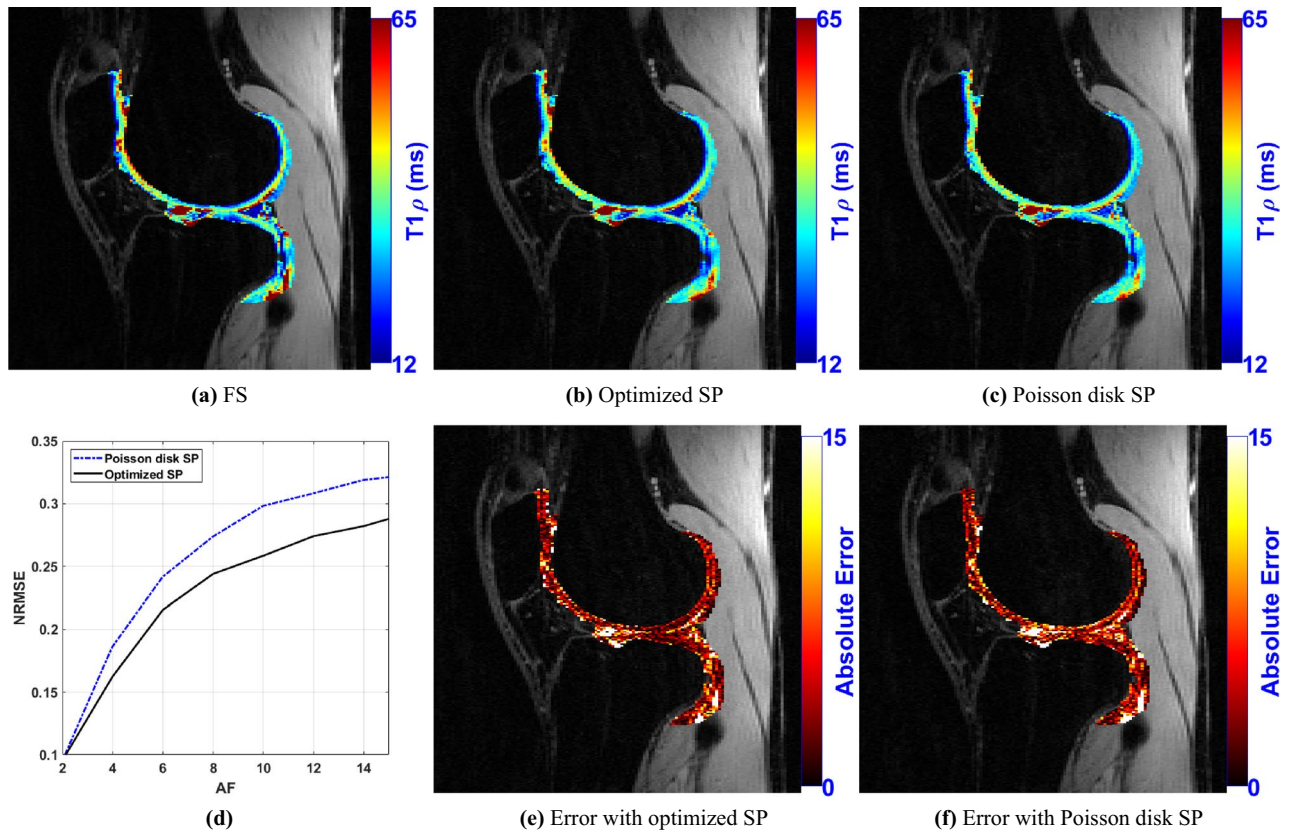
Figure 10. Comparison of $T_{1\rho}$ maps obtained with Poisson disk and optimized SP. Illustrative $T_{1\rho}$ maps obtained with fully sampled (FS) parallel MRI in (**a**), with optimized SP with AF = 4 in (**b**), and with Poisson disk SP with AF = 4 in (**c**). In (**d**), the NRMSE for different AF is shown. The point-wise error map between the $T_{1\rho}$ maps obtained between FS and optimized SP are shown in (**e**), and between FS and Poisson disk SP are shown in (**f**).
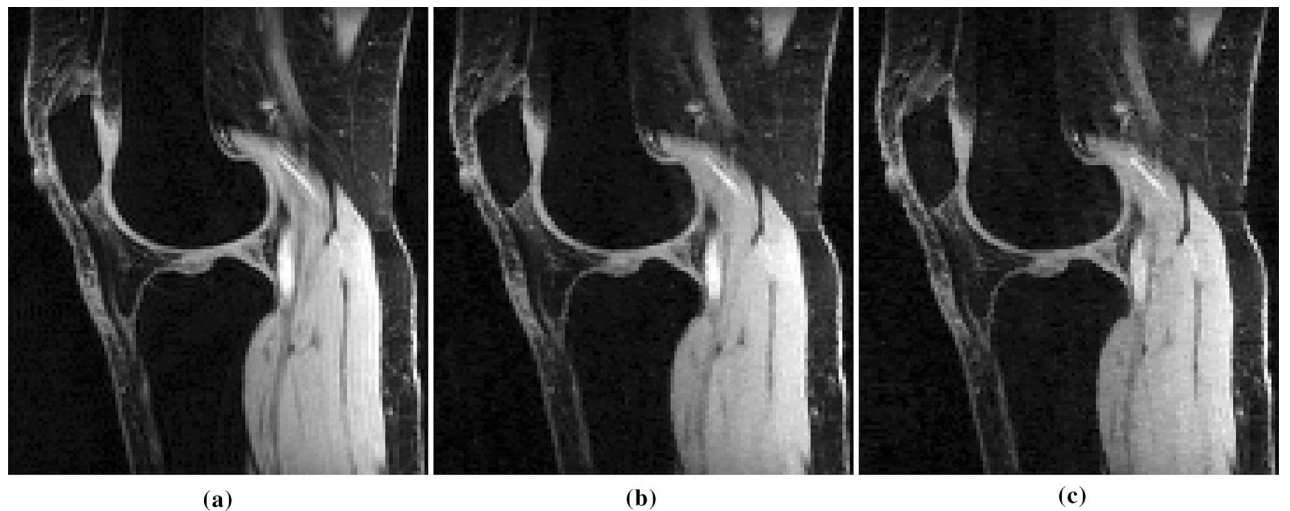

Figure 11. Comparison of (**a**) fully sampled parallel MRI (2 min 46 s scan time) with prospective accelerated parallel MRI at AF = 4 (42 s scan time) using (**b**) optimized SP and (**c**) Poisson disk SP.

considers measurements for specific anatomy and assumes a specific reconstruction method. Our experiments show that the optimized SPs are different for different reconstruction methods, suggesting that matching the sampling to the reconstruction method is important. The approach improves the acceleration factor and helps with finding the best SP for reconstruction methods in various applications of parallel MRI.

Data availability

Matlab codes and some sample data used for training and validation are available at https://cai2r.net/resources/software/data-driven-learning-sampling-pattern. Brain dataset is available at https://fastmri.med.nyu.edu/. Complete $T_{1\rho}$-brain and $T_{1\rho}$-knee datasets are available on request.

References

1. Bernstein, M., King, K. & Zhou, X. *Handbook of MRI Pulse Sequences* (Academic Press, 2004).
2. Liang, Z. P. & Lauterbur, P. C. *Principles of Magnetic Resonance Imaging: A Signal Processing Perspective* (IEEE Press, 2000).
3. Lustig, M., Donoho, D. L. & Pauly, J. M. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.* **58**, 1182–1195 (2007).
4. Trzasko, J. & Manduca, A. Highly undersampled magnetic resonance image reconstruction via homotopic ℓ_0-minimization. *IEEE Trans. Med. Imaging* **28**, 106–121 (2009).
5. Lustig, M., Donoho, D. L., Santos, J. M. & Pauly, J. M. Compressed sensing MRI. *IEEE Signal. Proc. Mag.* **25**, 72–82 (2008).
6. Jaspan, O. N., Fleysher, R. & Lipton, M. L. Compressed sensing MRI: A review of the clinical literature. *Br. J. Radiol.* **88**, 1056 (2015).
7. Shin, P. J. *et al.* Calibrationless parallel imaging reconstruction based on structured low-rank matrix completion. *Magn. Reson. Med.* **72**, 959–970 (2014).
8. Haldar, J. P. Low-rank modeling of local k-space neighborhoods (LORAKS) for constrained MRI. *IEEE Trans. Med. Imaging* **33**, 668–681 (2014).
9. Ongie, G. & Jacob, M. A fast algorithm for convolutional structured low-rank matrix recovery. *IEEE Trans. Comput. Imaging* **3**, 535–550 (2017).
10. Jin, K. H., Lee, D. & Ye, J. C. A general framework for compressed sensing and parallel MRI using annihilating filter based low-rank Hankel matrix. *IEEE Trans. Comput. Imaging* **2**, 480–495 (2016).
11. Haldar, J. P. & Zhuo, J. P-LORAKS: Low-rank modeling of local k-space neighborhoods with parallel imaging data. *Magn. Reson. Med.* **75**, 1499–1514 (2016).
12. Zibetti, M. V. W., Sharafi, A., Otazo, R. & Regatte, R. R. Accelerating 3D–$T1\rho$ mapping of cartilage using compressed sensing with different sparse and low rank models. *Magn. Reson. Med.* **80**, 1475–1491 (2018).
13. Zibetti, M. V. W., Herman, G. T. & Regatte, R. R. Data-driven design of the sampling pattern for compressed sensing and low rank reconstructions on parallel MRI of human knee joint. In *ISMRM Workshop on Data Sampling and Image Reconstruction* (2020).
14. Tsao, J. Ultrafast imaging: Principles, pitfalls, solutions, and applications. *J. Magn. Reson. Imaging* **32**, 252–266 (2010).
15. Ying, L. & Liang, Z. P. Parallel MRI using phased array coils. *IEEE Signal. Proc. Mag.* **27**, 90–98 (2010).
16. Pruessmann, K. P. Encoding and reconstruction in parallel MRI. *NMR Biomed.* **19**, 288–299 (2006).
17. Blaimer, M. *et al.* SMASH, SENSE, PILS, GRAPPA. *Top. Magn. Reson. Imaging* **15**, 223–236 (2004).
18. Wang, Y. Description of parallel imaging in MRI using multiple coils. *Magn. Reson. Med.* **44**, 495–499 (2000).
19. Wright, K. L., Hamilton, J. I., Griswold, M. A., Gulani, V. & Seiberlich, N. Non-Cartesian parallel imaging reconstruction. *J. Magn. Reson. Imaging* **40**, 1022–1040 (2014).
20. Feng, L. *et al.* Compressed sensing for body MRI. *J. Magn. Reson. Imaging* **45**, 966–987 (2017).
21. Lee, J. A. & Verleysen, M. *Nonlinear Dimensionality Reduction* (Springer Science + Business Media, LCC, 2007).
22. Elad, M., Figueiredo, M. A. T. & Ma, Y. On the role of sparse and redundant representations in image processing. *Proc. IEEE* **98**, 972–982 (2010).
23. Jacob, M., Mani, M. P. & Ye, J. C. Structured low-rank algorithms: Theory, magnetic resonance applications, and links to machine learning. *IEEE Signal. Proc. Mag.* **37**, 54–68 (2020).
24. Rubinstein, R., Bruckstein, A. M. & Elad, M. Dictionaries for sparse representation modeling. *Proc. IEEE* **98**, 1045–1057 (2010).
25. Knoll, F. *et al.* Deep-learning methods for parallel magnetic resonance imaging reconstruction: A survey of the current approaches, trends, and issues. *IEEE Signal. Proc. Mag.* **37**, 128–140 (2020).
26. Liang, D., Cheng, J., Ke, Z. & Ying, L. Deep magnetic resonance image reconstruction: Inverse problems meet neural networks. *IEEE Signal. Proc. Mag.* **37**, 141–151 (2020).
27. Candès, E. J. & Romberg, J. Sparsity and incoherence in compressive sampling. *Inverse Prob.* **23**, 969–985 (2007).
28. Donoho, D. L. Compressed sensing. *IEEE Trans. Inf. Theory* **52**, 1289–1306 (2006).
29. Haldar, J. P., Hernando, D. & Liang, Z. P. Compressed-sensing MRI with random encoding. *IEEE Trans. Med. Imaging* **30**, 893–903 (2011).
30. Unser, M. Sampling-50 years after Shannon. *Proc. IEEE* **88**, 569–587 (2000).
31. Candès, E. J. & Tao, T. Near optimal signal recovery from random projections: Universal encoding strategies?. *IEEE Trans. Inf. Theory* **52**, 5406–5425 (2006).
32. Candès, E. J., Romberg, J. K. & Tao, T. Stable signal recovery from incomplete and inaccurate measurements. *Commun. Pure Appl. Math.* **59**, 1207–1223 (2006).
33. Donoho, D. L., Elad, M. & Temlyakov, V. N. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inf. Theory* **52**, 6–18 (2006).
34. Adcock, B., Hansen, A. C., Poon, C. & Roman, B. Breaking the coherence barrier: A new theory for compressed sensing. *Forum Math. Sigma* **5**, e4 (2017).
35. Boyer, C., Bigot, J. & Weiss, P. Compressed sensing with structured sparsity and structured acquisition. *Appl. Comput. Harm. Anal.* **46**, 312–350 (2019).
36. Zijlstra, F., Viergever, M. A. & Seevinck, P. R. Evaluation of variable density and data-driven k-space undersampling for compressed sensing magnetic resonance imaging. *Invest. Radiol.* **51**, 410–419 (2016).
37. Boyer, C., Chauffert, N., Ciuciu, P., Kahn, J. & Weiss, P. On the generation of sampling schemes for magnetic resonance imaging. *SIAM J. Imaging Sci.* **9**, 2039–2072 (2016).
38. Cheng, J. Y. *et al.* Variable-density radial view-ordering and sampling for time-optimized 3D Cartesian imaging. In *ISMRM Workshop on Data Sampling and Image Reconstruction* (2013).
39. Ahmad, R. *et al.* Variable density incoherent spatiotemporal acquisition (VISTA) for highly accelerated cardiac MRI. *Magn. Reson. Med.* **74**, 1266–1278 (2015).
40. Wang, Z. & Arce, G. R. Variable density compressed image sampling. *IEEE Trans. Image Process.* **19**, 264–270 (2010).
41. Murphy, M. *et al.* Fast ℓ_1-spirit compressed sensing parallel imaging MRI: Scalable parallel implementation and clinically feasible runtime. *IEEE Trans. Med. Imaging* **31**, 1250–1262 (2012).
42. Kaldate, A., Patre, B. M., Harsh, R. & Verma, D. MR image reconstruction based on compressed sensing using Poisson sampling pattern. In *Second International Conference on Cognitive Computing and Information Processing (CCIP)* (IEEE, 2016).

43. Dwork, N. *et al.* Fast variable density Poisson-disc sample generation with directional variation for compressed sensing in MRI. *Magn. Reson. Imaging* **77**, 186–193 (2021).

44. Levine, E., Daniel, B., Vasanawala, S., Hargreaves, B. & Saranathan, M. 3D Cartesian MRI with compressed sensing and variable view sharing using complementary Poisson-disc sampling. *Mag. Reson. Med.* **77**, 1774–1785 (2017).

45. Knoll, F., Clason, C., Diwoky, C. & Stollberger, R. Adapted random sampling patterns for accelerated MRI. *Magn. Reson. Mater. Phys. Biol. Med.* **24**, 43–50 (2011).

46. Choi, J. & Kim, H. Implementation of time-efficient adaptive sampling function design for improved undersampled MRI reconstruction. *J. Magn. Reson.* **273**, 47–55 (2016).

47. Vellagoundar, J. & Machireddy, R. R. A robust adaptive sampling method for faster acquisition of MR images. *Magn. Reson. Imaging* **33**, 635–643 (2015).

48. Krishna, C. & Rajgopal, K. Adaptive variable density sampling based on Knapsack problem for fast MRI. In *IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)* 364–369 (IEEE, 2015).

49. Zhang, Y., Peterson, B. S., Ji, G. & Dong, Z. Energy preserved sampling for compressed sensing MRI. *Comput. Math. Methods Med.* **2014**, 1–12 (2014).

50. Kim, W., Zhou, Y., Lyu, J. & Ying, L. Conflict-cost based random sampling design for parallel MRI with low rank constraints. In *Compressive Sensing IV*, Vol. 9484 of *Compressive Sensing IV* (ed. Ahmad, F.) 94840 (2015).

51. Haldar, J. P. & Kim, D. OEDIPUS: An experiment design framework for sparsity-constrained MRI. *IEEE Trans. Med. Imaging* **38**, 1545–1558 (2019).

52. Seeger, M., Nickisch, H., Pohmann, R. & Schölkopf, B. Optimization of k-space trajectories for compressed sensing by Bayesian experimental design. *Magn. Reson. Med.* **63**, 116–126 (2010).

53. Zhao, B. *et al.* Optimal experiment design for magnetic resonance fingerprinting: Cramér–Rao bound meets spin dynamics. *IEEE Trans. Med. Imaging* **38**, 844–861 (2019).

54. Bouhrara, M. & Spencer, R. G. Fisher information and Cramér–Rao lower bound for experimental design in parallel imaging. *Magn. Reson. Med.* **79**, 3249–3255 (2018).

55. Gözcü, B. *et al.* Learning-based compressive MRI. *IEEE Trans. Med. Imaging* **37**, 1394–1406 (2018).

56. Gözcü, B., Sanchez, T. & Cevher, V. Rethinking sampling in parallel MRI: A data-driven approach. In *European Signal Processing Conference* (2019).

57. Sanchez, T. *et al.* Scalable learning-based sampling optimization for compressive dynamic MRI. In *IEEE International Conference on Acoustics, Speech and Signal Processing* 8584–8588 (IEEE, 2020).

58. Liu, D. D., Liang, D., Liu, X. & Zhang, Y. T. Under-sampling trajectory design for compressed sensing MRI. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society* 73–76 (IEEE, 2012).

59. Ravishankar, S. & Bresler, Y. Adaptive sampling design for compressed sensing MRI. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vol. 20, 3751–3755 (IEEE, 2011).

60. Bahadir, C. D., Wang, A. Q., Dalca, A. V. & Sabuncu, M. R. Deep-learning-based optimization of the under-sampling pattern in MRI. *IEEE Trans. Comput. Imaging* **6**, 1139–1152 (2020).

61. Aggarwal, H. K. & Jacob, M. J-MoDL: Joint model-based deep learning for optimized sampling and reconstruction. *IEEE J. Sel. Top. Signal Process.* **14**, 1151–1162 (2020).

62. Sherry, F. *et al.* Learning the sampling pattern for MRI. *IEEE Trans. Med. Imaging* **39**, 4310–4321 (2020).

63. Broughton, R., Coope, I., Renaud, P. & Tappenden, R. Determinant and exchange algorithms for observation subset selection. *IEEE Trans. Image Process.* **19**, 2437–2443 (2010).

64. Zhou, Z. H., Yu, Y. & Qian, C. *Evolutionary Learning: Advances in Theories And Algorithms* (Springer Singapore, 2019).

65. Couvreur, C. & Bresler, Y. On the optimality of the backward greedy algorithm for the subset selection problem. *SIAM J. Matrix Anal. Appl.* **21**, 797–808 (2000).

66. Qian, C., Yu, Y. & Zhou, Z. H. Subset selection by Pareto optimization. *Adv. Neural Inf. Process. Syst.* **2015**, 1774–1782 (2015).

67. Qian, C., Shi, J. C., Yu, Y., Tang, K. & Zhou, Z. H. Subset selection under noise. *Adv. Neural Inf. Process. Syst.* **2017**, 3561–3571 (2017).

68. Wen, B., Ravishankar, S., Pfister, L. & Bresler, Y. Transform learning for magnetic resonance image reconstruction: From model-based learning to building neural networks. *IEEE Signal. Proc. Mag.* **37**, 41–53 (2020).

69. Wang, G., Luo, T., Nielsen, J. F., Noll, D. C. & Fessler, J. A. B-spline Parameterized Joint Optimization of Reconstruction and K-space Trajectories (BJORK) for Accelerated 2D MRI. arXiv preprint arXiv:2101.11369 1–14 (2021).

70. Weiss, T. *et al.* PILOT: Physics-Informed Learned Optimized Trajectories for Accelerated MRI. arXiv preprint arXiv:1909.05773 1–12 (2019).

71. Bakker, T., Van Hoof, H. & Welling, M. Experimental design for MRI by greedy policy search. In *Advances in Neural Information Processing Systems*. arXiv:2010.16262 (2020).

72. Jin, K. H., Unser, M. & Yi, K. M. Self-Supervised Deep Active Accelerated MRI. arXiv preprint 1901.04547 1–13 (2019). arXiv:1901.04547.

73. Pineda, L., Basu, S., Romero, A., Calandra, R. & Drozdzal, M. Active MR k-space Sampling with Reinforcement Learning. *Lect. Notes Comput. Sci.* **12262 LNCS**, 23–33 (2020). arXiv:2007.10469.

74. Zhang, Z. *et al.* Reducing Uncertainty in Undersampled MRI Reconstruction with Active Acquisition. In *IEEE Conference on Computer Vision and Pattern Recognition* 2049–2058 (IEEE, 2019). arXiv:1902.03051.

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

76. Zibetti, M. V. W., Baboli, R., Chang, G., Otazo, R. & Regatte, R. R. Rapid compositional mapping of knee cartilage with compressed sensing MRI. *J. Magn. Reson. Imaging* **48**, 1185–1198 (2018).

77. Uecker, M. *et al.* ESPIRiT-an eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA. *Magn. Reson. Med.* **71**, 990–1001 (2014).

78. Walsh, D. O., Gmitro, A. F. & Marcellin, M. W. Adaptive reconstruction of phased array MR imagery. *Magn. Reson. Med.* **43**, 682–690 (2000).

79. Roemer, P. B., Edelstein, W. A., Hayes, C. E., Souza, S. P. & Mueller, O. M. The NMR phased array. *Magn. Reson. Med.* **16**, 192–225 (1990).

80. Liu, B., Zou, Y. M. & Ying, L. SparseSENSE: Application of compressed sensing in parallel MRI. In *International Conference on Technology and Applications in Biomedicine*, Vol. 2, 127–130 (IEEE, 2008).

81. Liang, D., Liu, B., Wang, J. & Ying, L. Accelerating SENSE using compressed sensing. *Magn. Reson. Med.* **62**, 1574–1584 (2009).

82. Haldar, J. P. Autocalibrated LORAKS for fast constrained MRI reconstruction. In *IEEE International Symposium on Biomedical Imaging* 910–913 (IEEE, 2015).

83. Liang, Z. P. Spatiotemporal imaging with partially separable functions. In *IEEE International Symposium on Biomedical Imaging*, Vol. 2, 988–991 (IEEE, 2007).

84. Elad, M., Milanfar, P. & Rubinstein, R. Analysis versus synthesis in signal priors. *Inverse Prob.* **23**, 947–968 (2007).

85. Doneva, M. *et al.* Compressed sensing reconstruction for magnetic resonance parameter mapping. *Magn. Reson. Med.* **64**, 1114–1120 (2010).

86. Zibetti, M. V. W., Helou, E. S., Sharafi, A. & Regatte, R. R. Fast multicomponent 3D-T1ρ relaxometry. *NMR Biomed.* e4318 (2020).

87. Zibetti, M. V. W., Helou, E. S., Regatte, R. R. & Herman, G. T. Monotone FISTA with variable acceleration for compressed sensing magnetic resonance imaging. *IEEE Trans. Comput. Imaging* **5**, 109–119 (2019).
88. Glover, F. & Laguna, M. Tabu search. In *Handbook of Applied Optimization* (eds Pardalos, P. M. & Resende, M. G. C.) 194–209 (Oxford University Press, 2002).
89. Knoll, F. *et al.* fastMRI: A publicly available raw k-Space and DICOM dataset of knee images for accelerated MR image reconstruction using machine learning. *Radiology* **2**, e190007 (2020).
90. Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**, 600–612 (2004).
91. Zhu, Y. *et al.* Bio-SCOPE: fast biexponential $T1\rho$ mapping of the brain using signal-compensated low-rank plus sparse matrix decomposition. *Magn. Reson. Med.* **83**, 2092–2106 (2020).

Acknowledgements

Author contributions

M.V.W.Z. developed the main algorithm, coded, and performed the main experiments. G.T.H. revised the main algorithm, provided input to the data analysis and manuscript. R.R.R. conceived and designed the study and provided input to the data analysis, and manuscript. All authors reviewed and contributed to the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.V.W.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.