SCALABLE LEARNING-BASED SAMPLING OPTIMIZATION FOR COMPRESSIVE DYNAMIC MRI

Thomas Sanchez[1], Baran Gözcü[1], Ruud B. van Heeswijk[2], Armin Eftekhari[3], Efe Ilıcak[4], Tolga Çukur[4], and Volkan Cevher[1]

[1]EPFL, Switzerland [2]CHUV, Switzerland [3]Umeå University, Sweden [4]Bilkent University, Turkey

ABSTRACT

Compressed sensing applied to magnetic resonance imaging (MRI) allows to reduce the scanning time by enabling images to be reconstructed from highly undersampled data. In this paper, we tackle the problem of designing a sampling mask for an arbitrary reconstruction method and a limited acquisition budget. Namely, we look for an optimal probability distribution from which a mask with a fixed cardinality is drawn. We demonstrate that this problem admits a compactly supported solution, which leads to a deterministic optimal sampling mask. We then propose a stochastic greedy algorithm that **(i)** provides an approximate solution to this problem, and **(ii)** resolves the scaling issues of [1, 2]. We validate its performance on *in vivo* dynamic MRI with retrospective undersampling, showing that our method preserves the performance of [1, 2] while reducing the computational burden by a factor close to 200. Our implementation is available at https://github.com/t-sanchez/stochasticGreedyMRI.

Index Terms— Magnetic resonance imaging, compressive sensing (CS), learning-based sampling.

1. INTRODUCTION

Dynamic Magnetic Resonance Imaging (dMRI) is a powerful tool in medical imaging, which allows for non-invasive monitoring of tissues over time. A main challenge to the quality of dMRI examinations is the inefficiency of data acquisition that limits temporal and spatial resolutions. In the presence of moving tissues, such as in cardiac MRI, the trade-off between spatial and temporal resolution is further complicated by the need to perform breath-holds to minimize motion artifacts [3].

In the last decade, the rise of Compressed Sensing (CS) has significantly contributed to overcoming these problems. CS allows for a successful reconstruction from undersampled measurements, provided that they are incoherent [4, 5] and that the data can be sparsely represented in some domain. In dMRI, samples are acquired in the k-t space (spatial frequency and time domain), and can be sparsely represented in the x-f domain (image and temporal Fourier transform domain). Many algorithms have exploited this framework with great success (see [6–14] and the references therein).

While CS theory mostly focuses on fully random measurements [15], the practical implementations have generally exploited *random variable-density sampling*, based on drawing random samples from a parametric distribution (typically polynomial or Gaussian) which reasonably imitates the energy distribution in the k-t space [16, 17]. While all these approaches allow to quickly design masks which yield a great improvement over fully random sampling, prescribed by the theory of CS, they **(i)** remain largely heuristic; **(ii)** ignore the anatomy of interest; **(iii)** ignore the reconstruction algorithm; **(iv)** require careful tuning of their various parameters, and **(v)** do not necessarily use a fixed number of readouts per frame.

In the present work, we show that the problem of finding an optimal mask sampling distribution which contains n out of p possible locations admits a solution compactly supported on n elements. This demonstrates that our previously proposed framework in [1, 2], which searches for an approximately optimal sampling mask, is in fact looking for a solution to the more general problem of finding an optimal measurement distribution. In addition, we propose a *scalable learning-based framework* for dMRI. Our proposed stochastic greedy method preserves the performance of [1, 2] while reducing the computational burden by a factor close to 200.

Numerical evidence shows that our framework can successfully find sampling patterns for a broad range of decoders, from k-t FOCUSS [7] to ALOHA [13], outperforming state-of-the-art model-based sampling methods over nearly all sampling rates considered.

2. THEORY

2.1. Signal Acquisition

In the compressed sensing (CS) problem [5], one desires to retrieve a signal that is known to be sparse in some basis using only a small number of linear measurements. In the case of dynamic MRI, we consider a signal $\mathbf{x} \in \mathbb{C}^p = \mathbb{C}^{N^2T}$ (i.e. a vectorized video of size $N \times N$ with T frames), and the subsampled Fourier measurements are

$$\mathbf{b} = \mathbf{P}_\Omega \mathbf{\Psi} \mathbf{x} + \mathbf{w} \tag{1}$$

where $\mathbf{\Psi} \in \mathbb{C}^p$ is the spatial Fourier transform operator applied to the vectorized signal, $\mathbf{P}_\Omega : \mathbb{C}^p \to \mathbb{C}^n$ is a subsam-

This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement n° 725594 - time-data), from Hasler Foundation Program: Cyber Human Systems (project number 16066) and from the Office of Naval Research (ONR) (grant n° N62909-17-1-2111).

pling operator that selects the rows of $\mathbf{\Psi}$ according to the indices in the set Ω with $|\Omega| = n$ and $n \ll p$. We refer to Ω as *sampling pattern* or *mask*. We assume the signal $\mathbf{x}$ to be sparse in the basis $\mathbf{\Phi}$, which typically is a temporal Fourier transform across frames. Given the samples $\mathbf{b}$, along with Ω, a *reconstruction algorithm* or *decoder* g forms an estimate $\hat{\mathbf{x}} = g(\mathbf{b}, \Omega)$ of $\mathbf{x}$.

The quality of the reconstruction is then evaluated using a *performance metric* $\eta(\mathbf{x}, \hat{\mathbf{x}})$, which could typically include Peak Signal-to-Noise Ratio (PSNR), the negative Mean Square Error (MSE), or the Structural Similarity Index Measure (SSIM) [18].

2.2. Sampling mask design

We model the mask designing process as finding a probability mass function (PMF) $f \in S^{p-1}$, where $S^{p-1} := \{f \in [0,1]^p : \sum_{i=1}^p f_i = 1\}$ is the standard simplex in $\mathbb{R}^p$. f assigns to each location i in the k-space a probability f_i to be acquired. The mask is then constructed by drawing without replacement from f until the cardinality constraint $|\Omega| = n$ is met. The problem of finding the optimal sampling distribution is subsequently formulated as

$$\max_{f \in S^{p-1}} \eta(f), \qquad \eta(f) := \mathbb{E}_{\substack{\Omega(f,n) \\ \mathbf{x} \sim \mathcal{P}_\mathbf{x}}} \left[\eta\left(\mathbf{x}, \hat{\mathbf{x}}\left(\Omega, \mathbf{x}\right)\right)\right], \quad (2)$$

where the index set $\Omega \subset [p]$ is generated from f and $[p] := \{1, \dots, p\}$. This problem corresponds to finding the probability distribution f that maximizes the expected performance metric with respect to the data $\mathcal{P}_\mathbf{x}$ and the masks drawn from this distribution. To ease the notation, we will use $\eta\left(\mathbf{x}, \hat{\mathbf{x}}\left(\Omega, \mathbf{x}\right)\right) \equiv \eta\left(\mathbf{x}; \Omega\right)$.

In practice, we do not have access to $\mathbb{E}_{\mathcal{P}_\mathbf{x}}\left[\eta(\mathbf{x}; \Omega)\right]$ and instead have at hand the training images $\{\mathbf{x}_i\}_{i=1}^m$ drawn independently from $\mathcal{P}_\mathbf{x}$. We therefore maximize the empirical perfromance by solving

$$\max_{f \in S^{p-1}} \eta_m(f), \quad \eta_m(f) := \frac{1}{m} \sum_{i=1}^m \mathbb{E}_{\Omega(f,n)} \left[\eta(\Omega, \mathbf{x}_i)\right]. \quad (3)$$

Given that Problem (3) looks for masks that are constructed by sampling n times without replacement from f, the following holds.

Proposition 1. *There exists a maximizer of Problem* (3) *that is supported on an index set of size at most* n.

Proof. Let the distribution $\widehat{f}_n$ be a maximizer of Problem (3). We are interested in finding the support of $\widehat{f}_n$. Because $\sum_{|\Omega|=n} \Pr[\Omega] = 1$, note that

$$\begin{aligned}
\max_{f \in S^{p-1}} \eta_m(f) &:= \max_{f \in S^{p-1}} \sum_{|\Omega|=n} \frac{1}{m} \sum\nolimits_{i=1}^m \eta(\mathbf{x}_i; \Omega) \cdot \Pr[\Omega | f] \\
&\leq \max_{f \in S^{p-1}} \max_{|\Omega|=n} \frac{1}{m} \sum\nolimits_{i=1}^m \eta(\mathbf{x}_i; \Omega) \\
&= \max_{|\Omega|=n} \frac{1}{m} \sum\nolimits_{i=1}^m \eta(\mathbf{x}_i; \Omega). \qquad (4)
\end{aligned}$$

Let $\widehat{\Omega}_n$ be an index set of size n that maximizes the last line above. The above holds with equality when $\Pr[\widehat{\Omega}_n] = 1$ and $\Pr[\Omega] = 0$ for $\Omega \neq \widehat{\Omega}_n$ and $f = \widehat{f}_n$. This in turn happens when $\widehat{f}_n$ is supported on $\widehat{\Omega}$. That is, there exists a maximizer of Problem (3) that is supported on an index set of size n. □

While this observation does not indicate how to find this maximizer, it nonetheless allows us to further simplify Problem (3). More specifically, the observation that a distribution $\widehat{f}_n$ has a compact support of size n implies the following:

Proposition 2.

$$\textit{Problem (3)} \equiv \max_{|\Omega|=n} \frac{1}{m} \sum_{i=1}^m \eta(\mathbf{x}_i; \Omega) \quad (5)$$

Proof. Proposition 1 tells us that a solution of Problem (3) is supported on a set of size at most n, which implies

$$\text{Problem (3)} \equiv \max_{f \in S^{p-1}, |\text{supp}(f)|=n} \eta_m(f). \quad (6)$$

That is, we only need to search over compactly supported distributions f. Let S_Γ denote the standard simplex on a support $\Gamma \subset [p]$. It holds that

$$\begin{aligned}
\text{Problem (6)} &\equiv \max_{|\Gamma|=n} \max_{f \in S_\Gamma} \eta_m(f) \\
&= \max_{|\Gamma|=n} \max_{f \in S_\Gamma} \frac{1}{m} \sum\nolimits_{i=1}^m \eta(\mathbf{x}_i; \Gamma) \cdot \Pr[\Gamma | f] \\
&= \max_{|\Gamma|=n} \max_{f \in S_\Gamma} \frac{1}{m} \sum\nolimits_{i=1}^m \eta(\mathbf{x}_i; \Gamma) \\
&= \max_{|\Gamma|=n} \frac{1}{m} \sum\nolimits_{i=1}^m \eta(\mathbf{x}_i; \Gamma). \qquad (7)
\end{aligned}$$

To obtain the second and third equalities, one observes that all masks have a common support Γ with n elements, i.e. $f \in S_\Gamma$ allows only for a single mask Ω with n elements, namely $\Omega = \Gamma$. □

The framework of Problem (3) captures most variable-density based approaches of the literature that are defined in a data-driven fashion [19–25], and Proposition 2 shows that Problem (7), that we tackled in [1, 2] and develop here, also aims at solving the *same* problem as these probabilistic approaches. Note that while the present theory considered sampling *points* in the Fourier space, it is readily applicable to the Cartesian case, where full lines are added to the mask at once.

3. STOCHASTIC GREEDY MASK DESIGN

Aligned with the approach that we previously proposed in [1], we want to find an approximate solution to Problem (5) by leveraging a greedy algorithm. This is required by Problem (5) being inherently combinatorial. The previous greedy method of [1, 2] suffers from three main drawbacks: **(i)** it scales quadratically with the total number of lines, **(ii)** it scales linearly with the size of the dataset, and **(iii)** it does not construct mask with a fixed number of readouts by frame. While [2] partially deals with **(i)**, our proposed stochastic greedy approach addresses all three issues, while preserving the benefits of [1]. It notably still preserves the nestedness and ordering of the acquisition, where critical locations are acquired initially, and the mask built outputs a nested structure (i.e. the mask at 30% sampling rate includes all sampling locations of the mask at 20%).

Let us introduce the set $\mathcal{S}$ of all lines that can be acquired, which is a set of subsets of $\{1, \dots, p\}$. A feasible Cartesian mask takes the form $\Omega = \bigcup_{j=1}^{\ell} S_j, \quad S_j \in \mathcal{S}$, i.e. it consists of a union of lines. Both the greedy method of [1] and our stochastic method are detailed in Algorithm 1 below. Our stochastic greedy method (**SG-v2**) addresses the three main limitations of the greedy method of [1] (**G-v1**). The issue **(i)** is solved by picking uniformly at random at each iteration a batch possible lines $\mathcal{S}_{iter}$ of size k from a given frame $\mathcal{S}_t$, instead of considering the full set of possible lines $\mathcal{S}$ (line 3 in Alg. 1); **(ii)** is addressed by considering a fixed batch of training data $\mathcal{L}$ of size l instead of the whole training set of size m at each iteration (line 4 in Alg. 1); **(iii)** is solved by iterating through the lines to be added from each frame $\mathcal{S}_t$ sequentially (lines 1, 3 and 10 in Alg. 1). These improvements are inspired by the refinements done to the standard greedy algorithm in the field of submodular optimization [26], and allow to move the computational complexity from $\Theta\left(mr(NT)^2\right)$ to $\Theta\left(lrkNT\right)$, effectively speeding up the computation by a factor $\mathbf{\Theta}(\frac{\mathbf{m}}{\mathbf{l}}\frac{\mathbf{NT}}{\mathbf{k}})$. Our results show that this is achieved without sacrificing any reconstruction quality.

Algorithm 1 Greedy mask optimization algorithms for dMRI
(G) refers to the greedy algorithm [1]
(SG) refers to the **stochastic greedy** algorithm
(v1) algorithm iterated throughout the whole training set
(v2) algorithm iterated through batches of training examples

Input: Training data $\{\mathbf{x}\}_{i=1}^{m}$, recon. rule g, sampling set $\mathcal{S}$, max. cardinality n, samp. batch size k, train. batch size l
Output: Sampling pattern Ω

1: **(SG)** Initialize $t = 1$
2: **while** $|\Omega| \leq n$ **do**
3: $\begin{cases} \textbf{(G)} & \text{Pick } \mathcal{S}_{iter} = \mathcal{S} \\ \textbf{(SG)} & \text{Pick } \mathcal{S}_{iter} \subseteq \mathcal{S}_t \text{ at random, with } |\mathcal{S}_{iter}| = k \end{cases}$
4: $\begin{cases} \textbf{(v1)} & \text{Pick } \mathcal{L} = \{1, \dots, m\} \\ \textbf{(v2)} & \text{Pick } \mathcal{L} \subseteq \{1, \dots, m\}, \text{ with } |\mathcal{L}| = l \end{cases}$
5: **for** $S \in \mathcal{S}_{iter}$ such that $|\Omega \cup S| \leq \Gamma$ **do**
6: $\Omega' = \Omega \cup S$
7: For each $\ell \in \mathcal{L}$ set $\hat{\mathbf{x}}_\ell \leftarrow g(\Omega', \mathbf{P}_{\Omega'}\mathbf{\Psi}\mathbf{x}_\ell)$
8: $\eta(\Omega') \leftarrow \frac{1}{|\mathcal{L}|}\sum_{\ell \in \mathcal{L}} \eta(\mathbf{x}_\ell, \hat{\mathbf{x}}_\ell)$
9: $\Omega \leftarrow \Omega \cup S^*$, where $S^* = \underset{S:|\Omega \cup S| \leq n}{\operatorname{argmax}} \ \eta(\Omega \cup S)$
10: **(SG)** $t = (t \bmod T) + 1$
11: **return** Ω

4. NUMERICAL EXPERIMENTS

4.1. Implementation details

Reconstruction algorithms: We consider three reconstruction algorithms, namely k-t *FOCUSS* (KTF) [7], and *ALOHA* [13]. Their parameters were selected to maintain a good empirical performance across all sampling rates considered.

Mask selection baselines:

- *Coherence-VD* [16]: We consider a random *variable-density* sampling mask with Gaussian density and optimize its parameters to minimize coherence.
- *LB-VD* [1, 2]: Instead of minimizing the coherence as in *Coherence-VD*, we perform a grid search on the parameters using the training set to optimize reconstruction according to the same performance metric as our method.

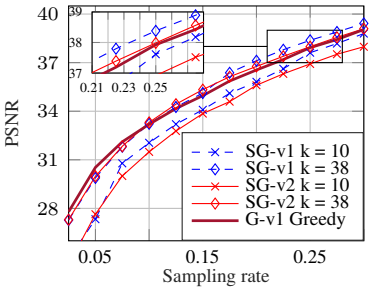

Fig. 1: PSNR as a function of the sampling rate for KTF, comparing the different reconstruction methods as well as the effect of the batch size on the quality of the reconstruction for SG.

Data sets: Our dynamic data were acquired in seven adult volunteers with a balanced steady-state free precession (bSSFP) pulse sequence on a whole-body Siemens 3T scanner using a 34-element matrix coil array. Several short-axis cine images were obtained during a breath-hold scan. Fully sampled Cartesian data were acquired using a 256×256 grid with 25 frames, then combined and cropped to a $152 \times 152 \times 17$ single coil image. The details of the parameters used are provided in the supplementary material [27]. In the experiments, we used three volumes for training and four for testing.

4.2. Comparison of greedy algorithms

We first compare the performance of G-v1 with SG-v1 and SG-v2, and show the results on Figure 1. We are specifically interested in determining the sensitivity of our algorithm to the sampling batch size k and training batch size l (for SG-v2, we use $l = 1$ unless stated differently). We see that using a small batch size k (e.g. 10) yields a drop in performance, while $k = 38$ even improves performance compared to G-v1, with respectively 60 times less computation for SG-v1 and 180 less computations for SG-v2. One should also note that using a batch of training images (SG-v2) does not reduce the performance compared to SG-v1, while largely reducing computations. Also, additional results (in the supplementary material [27]) show that using larger batches yields similar results as for $k = 38$. The fact that the performance of SG-v2 with $k = 38$ outperforms G-v1 could be surprising, but originates in the lack of structure of the problem, where introducing noise in the computations through random batches of samples improves the overall performance of the method. In the sequel, we use $k = 38$ and $l = 1$ for SCG-v2.

4.3. Single coil results

The comparison to baselines is shown on Figures 2 and 3, where we see that the SG-v2 method yields masks that consistently improve the results compared to all variable-density methods used.

We notice in Figure 3 that comparing the reconstruction algorithms with VD methods do not allow for a faithful performance comparison of the reconstruction algorithms: the

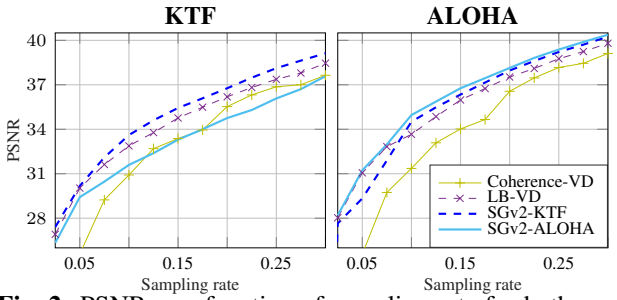

Fig. 2: PSNR as a function of sampling rate for both reconstruction algorithms considered, comparing the mask design methods considered, averaged on 4 images.

performance difference is very small between the reconstruction methods. In contrast, considering the reconstruction algorithm jointly with a sampling pattern optimized with our model-free approach makes the performance difference much more noticeable: ALOHA with its corresponding mask clearly outperforms KTF, and this conclusion could not be made by looking solely at reconstructions with VD-based masks. Note that extended results, along with multi-coil experiments, are available in our supplementary material [27].

4.4. Large scale static results

This last experiment shows the scalability of our method to very large datasets. We used the fastMRI dataset [28] consisting of knee volumes and trained the mask for reconstructing the 13 most central slices of size 320×320, which yielded a training set containing 12649 slices. For the sake of brevity, we only report computations performed using total variation (TV) minimization with NESTA [29]. For mask design, we used the SG-v2 method with $k = 80$ and $l = 20$ (2500 fewer computations compared to G-v1). The LB-VD method was trained using 80 representative slices and optimizing the parameters with a similar computational budget as SG-v2. The result on Figure 4 shows a uniform improvement of our method over the LB-VD approach.

5. DISCUSSION AND CONCLUSION

We presented a scalable sampling optimization method for dMRI, which largely addresses the scalability issues of [1, 2]. Reducing the resources used by G-v1 by as much as a 200 times was shown to have no negative impact on the quality of reconstruction achieved within our framework. Our method was demonstrated to successfully scale to very large datasets such as fastMRI [28], which the previous greedy method [1] could not achieve.

The masks obtained bring significant image quality improvements over the baselines. The results suggest that VD-based methods limit the performance of CS applied to MRI through their underlying model. They are consistently outperformed by our model-free and data-adaptive method on different *in vivo* datasets, across several decoders, field of views and resolutions. Our findings highlight that sampling design should not be considered in isolation from data and reconstruction algorithm, as using a mask that is not specifically optimized can considerably hinder the performance of the algorithm.

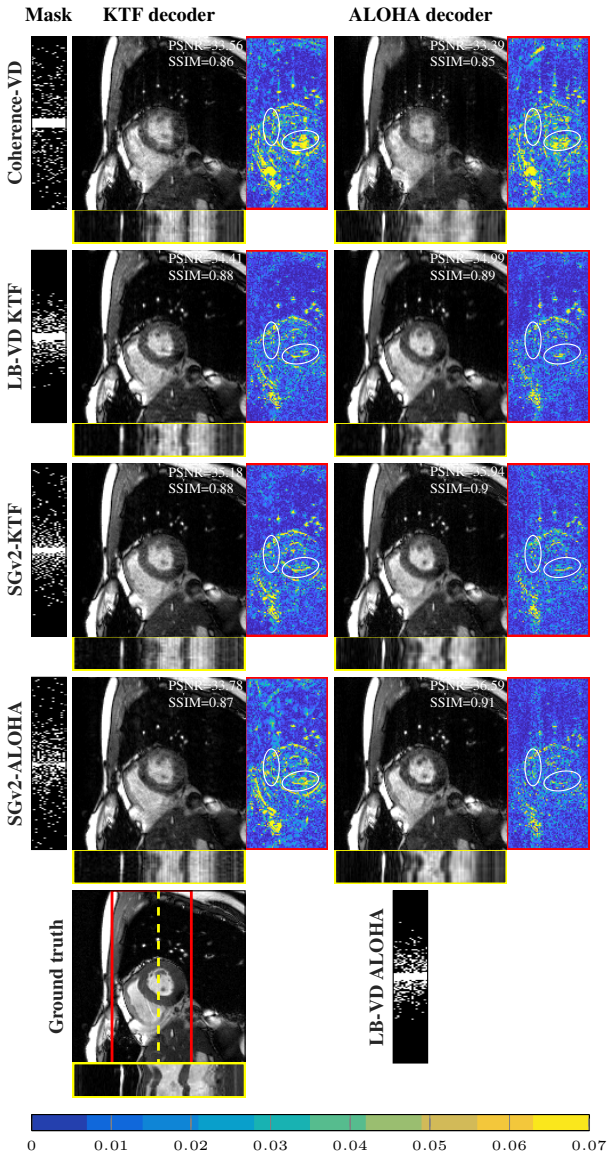

Fig. 3: Comparison of the different reconstruction masks and decoders, for a sampling rate of 15% on a single sample with its PSNR/SSIM performances.

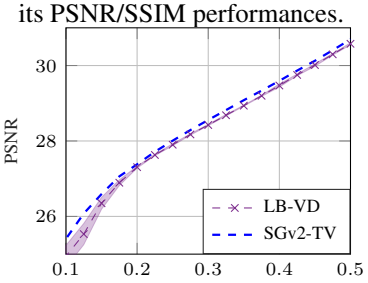

Fig. 4: PSNR as a function of the sampling rate for TV, averaged on the 13 most central slices of the fastMRI validation set [28] (2587 slices). SGv2 outperforms LB-VD over all sampling rates.

More importantly, our theoretical results show that the generic non-convex Problem (3) aiming at finding a probability mass function under a cardinality constraint from which a mask is subsequently sampled, is equivalent to the discrete Problem (7) of looking for the support of this PMF. This connection opens the door to rigorously leveraging techniques from combinatorial optimization for the problem of designing optimal, data-driven sampling masks for MRI.

6. REFERENCES

[1] B. Gözcü, R. K. Mahabadi, Y.-H. Li, E. Ilıcak, T. Çukur, J. Scarlett, and V. Cevher, "Learning-based compressive MRI," *IEEE Transactions on Medical Imaging*, 2018.

[2] B. Gözcü, T. Sanchez, and V. Cevher, "Rethinking sampling in parallel MRI: A data-driven approach," in *27th European Signal Processing Conference*, 2019.

[3] M. Saeed, T. A. Van, R. Krug, S. W. Hetts, and M. W. Wilson, "Cardiac MR imaging: current status and future direction," *Cardiovascular diagnosis and therapy*, vol. 5, no. 4, p. 290, 2015.

[4] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on pure and applied mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[5] D. L. Donoho, "Compressed sensing," *IEEE transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[6] M. Lustig, J. M. Santos, D. L. Donoho, and J. M. Pauly, "$k-t$ SPARSE: High frame rate dynamic MRI exploiting spatio-temporal sparsity," in *Proc. of the 13th Annual Meeting of ISMRM, Seattle*, vol. 2420, 2006.

[7] H. Jung, K. Sung, K. S. Nayak, E. Y. Kim, and J. C. Ye, "k-t FOCUSS: A general compressed sensing framework for high resolution dynamic MRI," *Magn. Reson. Med.*, vol. 61, no. 1, pp. 103–116, 2009.

[8] R. Otazo, D. Kim, L. Axel, and D. K. Sodickson, "Combination of compressed sensing and parallel imaging for highly accelerated first-pass cardiac perfusion MRI," *Magnetic Resonance in Medicine*, vol. 64, no. 3, pp. 767–776, 2010.

[9] S. G. Lingala, Y. Hu, E. DiBella, and M. Jacob, "Accelerated dynamic MRI exploiting sparsity and low-rank structure: k-t SLR," *IEEE Transactions on Medical Imaging*, vol. 30, no. 5, pp. 1042–1054, 2011.

[10] L. Feng, R. Grimm, K. T. Block, H. Chandarana, S. Kim, J. Xu, L. Axel, D. K. Sodickson, and R. Otazo, "Golden-angle radial sparse parallel MRI: Combination of compressed sensing, parallel imaging, and golden-angle radial sampling for fast and flexible dynamic volumetric MRI," *Magnetic Resonance in Medicine*, vol. 72, no. 3, pp. 707–717, 2014.

[11] J. Caballero, A. N. Price, D. Rueckert, and J. V. Hajnal, "Dictionary learning and time sparsity for dynamic MR data reconstruction," *IEEE Transactions on Medical Imaging*, vol. 33, no. 4, pp. 979–994, 2014.

[12] R. Otazo, E. Candès, and D. K. Sodickson, "Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components," *MRM*, vol. 73, no. 3, pp. 1125–1136, 2015.

[13] K. H. Jin, D. Lee, and J. C. Ye, "A general framework for compressed sensing and parallel MRI using annihilating filter based low-rank Hankel matrix," *IEEE Transactions on Computational Imaging*, vol. 2, no. 4, pp. 480–495, 2016.

[14] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert, "A deep cascade of convolutional neural networks for dynamic MR image reconstruction," *IEEE Transactions on Medical Imaging*, vol. 37, no. 2, pp. 491–503, 2018.

[15] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans, on Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[16] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[17] H. Jung, J. C. Ye, and E. Y. Kim, "Improved k–t BLAST and k–t SENSE using FOCUSS," *Physics in medicine and biology*, vol. 52, no. 11, p. 3201, 2007.

[18] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[19] M. Seeger, H. Nickisch, R. Pohmann, and B. Schölkopf, "Optimization of k-space trajectories for compressed sensing by bayesian experimental design," *Magn. Reson. Med.*, vol. 63, no. 1, pp. 116–126, 2010.

[20] S. Ravishankar and Y. Bresler, "Adaptive sampling design for compressed sensing MRI," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*. IEEE, 2011, pp. 3751–3755.

[21] J. Vellagoundar and R. R. Machireddy, "A robust adaptive sampling method for faster acquisition of MR images," *Magnetic resonance imaging*, vol. 33, no. 5, pp. 635–643, 2015.

[22] L. Weizman, Y. C. Eldar, and D. Ben Bashat, "Compressed sensing for longitudinal MRI: An adaptive-weighted approach," *Medical physics*, vol. 42, no. 9, pp. 5195–5208, 2015.

[23] J. P. Haldar and D. Kim, "Oedipus: An experiment design framework for sparsity-constrained MRI," *IEEE transactions on medical imaging*, 2019.

[24] C. D. Bahadir, A. V. Dalca, and M. R. Sabuncu, "Learning-based optimization of the under-sampling pattern in MRI," in *International Conference on Information Processing in Medical Imaging*. Springer, 2019, pp. 780–792.

[25] F. Sherry, M. Benning, J. C. D. l. Reyes, M. J. Graves, G. Maierhofer, G. Williams, C.-B. Schönlieb, and M. J. Ehrhardt, "Learning the sampling pattern for MRI," *arXiv preprint arXiv:1906.08754*, 2019.

[26] B. Mirzasoleiman, A. Badanidiyuru, A. Karbasi, J. Vondrák, and A. Krause, "Lazier than lazy greedy." in *AAAI*, 2015, pp. 1812–1818.

[27] T. Sanchez, B. Gözcü, A. Eftekhari, R. B. van Heeswijk, E. Ilıcak, T. Çukur, and V. Cevher, "Scalable learning-based compressive dynamic MRI," *arXiv preprint arXiv:1902.00386*.

[28] J. Zbontar, F. Knoll, A. Sriram, M. J. Muckley, M. Bruno, A. Defazio, M. Parente, K. J. Geras, J. Katsnelson, H. Chandarana *et al.*, "fastMRI: An open dataset and benchmarks for accelerated MRI," *arXiv preprint arXiv:1811.08839*, 2018.

[29] S. Becker, J. Bobin, and E. J. Candès, "Nesta: A fast and accurate first-order method for sparse recovery," *SIAM Journal on Imaging Sciences*, vol. 4, no. 1, pp. 1–39, 2011.