CoverBLIP: accelerated and scalable iterative matched-filtering for Magnetic Resonance Fingerprint reconstruction

CoverBLIP: accelerated and scalable iterative matched-filtering for Magnetic Resonance Fingerprint reconstruction

Mohammad Golbabaee[1], Zhouye Chen[2], Yves Wiaux[3], and Mike Davies[4]

[1]Computer Science Department, University of Bath, UK, [2]Philips Research China, [3]School of Engineering and Physical Sciences, Heriot-Watt University, UK, [4]School of Engineering, University of Edinburgh, UK

E-mail: [1]m.golbabaee@bath.ac.uk

Abstract. Current popular methods for Magnetic Resonance Fingerprint (MRF) recovery are bottlenecked by the heavy computations of a matched-filtering step due to the growing size and complexity of the fingerprint dictionaries in multi-parametric quantitative MRI applications. We address this shortcoming by arranging dictionary atoms in the form of cover tree structures and adopt the corresponding fast approximate nearest neighbour searches to accelerate matched-filtering. For datasets belonging to smooth low-dimensional manifolds cover trees offer search complexities logarithmic in terms of data population. With this motivation we propose an iterative reconstruction algorithm, named CoverBLIP, to address large-size MRF problems where the fingerprint dictionary i.e. discrete manifold of Bloch responses, encodes several intrinsic NMR parameters. We study different forms of convergence for this algorithm and we show that provided with a notion of embedding, the inexact and non-convex iterations of CoverBLIP linearly convergence toward a near-global solution with the same order of accuracy as using exact brute-force searches. Our further examinations on both synthetic and real-world datasets and using different sampling strategies, indicates between 2 to 3 orders of magnitude reduction in total search computations. Cover trees are robust against the curse-of-dimensionality and therefore CoverBLIP provides a notion of scalability—a consistent gain in time-accuracy performance—for searching high-dimensional atoms which may not be easily preprocessed (i.e. for dimensionality reduction) due to the increasing degrees of non-linearities appearing in the emerging multi-parametric MRF dictionaries.

Keywords: Magnetic resonance fingerprinting, approximate projected gradient, cover trees, compressed sensing.[‡]

1. Introduction

Quantitative Magnetic Resonance Imaging (Q-MRI) provides a powerful tool for measuring various intrinsic NMR properties of tissues such as the $T1$, $T2$ and $T2^*$ relaxation times, field inhomogeneity, diffusion and perfusion [3]. As opposed to

[‡] Preliminary results related to this work were appeared as an IEEE MLSP conference paper [1] and an ISMRM abstract [2].

mainstream qualitative assessments these *absolute* physical quantities can be used for tissue or pathology identification independent of the scanner or scanning sequences. Despite being the long-standing goal of the MRI community, current quantitative approaches are extremely time-inefficient and for this reason not clinically applicable. The long process of acquiring multiple fully-sampled images for estimating each parameter brings serious limitation to the conventional Q-MRI approaches (e.g. [4, 5, 6, 7]) to apply within a reasonable time and with an acceptable signal-to-noise ratio (SNR) and resolution.

Recently Magnetic Resonance Fingerprinting (MRF) has emerged to address this short-coming and significantly accelerate the acquisition time of Q-MRI [8]. Three key principles are behind this new paradigm: i) applying *one* excitation sequence (i.e. in one acquisition run) that simultaneously encodes many quantitative parameters of interest, ii) incorporating more complicated (and sometimes random) but shorter excitation patterns than those used in conventional Q-MRI schemes, and finally iii) significant under-sampling of the k-space data at each temporal frame. The aggressively short acquisition times used in this framework, on the other hand, introduce several algorithmic challenges at the parameter estimation stage of the *MRF reconstruction problem*. Common approaches adopt a physical model to disambiguate the lack of sufficient spatio-temporal measurements in such a highly ill-posed inverse problem. This model is often not analytic and requires solving Bloch differential equations [9]. The MRF framework proposes to discretize the parameter space and exhaustively simulate a large dictionary of magnetic responses (fingerprints) to be used for *matched-filtering* in many model-based reconstruction routines (see e.g. [8, 10, 11, 12, 13, 14]). As occurs to any multi-parametric manifold enumeration, the main drawback of such approach is the size of this dictionary which grows exponentially in terms of the number of parameters and their quantization resolution. This brings a serious (scalability) limitation to the current popular schemes to be applicable in the emerging multi-parametric MRF problems [15, 16, 17, 18, 19, 20], as the computational complexity of exact matched-filtering using brute-force searches grows linearly with the dictionary size.

To address this shortcoming we propose an iterative reconstruction method with *inexact* updates dubbed as *Cover BLoch response Iterative Projection* (CoverBLIP). Our algorithm accelerates matched-filtering steps by replacing iterative brute-force searches with fast Approximate Nearest Neighbour Searches (ANNS) based on *cover tree* structures constructed off-line for a given MRF dictionary. For datasets living on smooth manifolds with low intrinsic dimension (e.g. a constant number of NMR characteristics) cover tree approximate searches are shown to have *logarithmic* complexity in terms of data populations [21]. Under an embedding assumption similar to the restricted isometry property in compressed sensing theory [22, 23, 24, 25, 26], we show that CoverBLIP iterations are able to correct the inexact updates and achieve a linear global convergence i.e. stable signal recovery. We also introduce an adaptive step-size scheme that guarantees (local) monotone convergence of CoverBLIP in general cases e.g. when the embedding assumption does not hold. The results provided in

this part apply beyond the customized MRF problem considered in this paper. We examine the reconstruction time-accuracy of the proposed method on both synthetic and *in-vivo* MRF datasets with different excitation sequences and k-space sampling patterns. Our experimental results indicate superiority of CoverBLIP compared to other tested baselines. Notably, CoverBLIP achieves 2-3 orders of magnitude acceleration in conducting matched-filtering while maintaining a similar accuracy as compared to using exact iterations with brute-force searches. Unlike non-scalable fast search algorithms such as KD-trees, we show that CoverBLIP maintains this superior performance when no dimensionality-reduction preprocessing is used. This feature of robustness against the high-dimensionality of search spaces makes CoverBLIP a well-suited candidate to tackle multi-parametric MRF applications with increased non-linear dynamic complexity, where applying common subspace compression preprocessing becomes prohibitive for their unfavourable compromise in the final estimation accuracy.

2. Related works

The seminal paper of Ma *et al.* [8] proposed a non-iterative Template Matching (TM) approach which consists of Fourier back-projections for all temporal slices followed by a per-voxel dictionary matching step. Adopted from *compressed sensing* literature Davies *et al.* [12] proposed an iterative extension to this framework and showed that repeated TM applications in the form of Iterative Projected Gradients (IPG) can highly improve parameter estimation in short acquisition sequences and low sampling/SNR regimes. The large size of the MRF dictionary is however a big challenge for the runtime of matched-filtering step(s) based on brute-force searches in both approaches. To address this issue, it has been proposed to reduce the (temporal) ambient search dimension using a few SVD bases of the MRF dictionary [11]. The main drawback of this approach is that the MRF dictionaries contain highly non-linear structures (a low-dimensional manifold of solutions of the Bloch equations) and therefore applying a linear subspace compression trades-off the computation time against the final accuracy of the reconstructed parameters. For instance, the Steady State Precession (FISP) sequence [10] encoding two NMR parameters requires 20 principal components to represent the search space within a reasonable accuracy, whereas an Inversion Recovery Balanced SSFP (IR-BSSFP) dictionary encoding three parameters requires 200 components [11]. One can imagine with the rise in applications encoding a larger number of parameters associated with the non-linear dynamics such as $T2^*$, perfusion, diffusion and microvascular properties, etc [15, 16, 17, 18, 19, 20] this issue will get worse i.e. an exponential growth in the dictionary size without a good low-dimensional subspace representation.

Another approach incorporates hierarchical clustering to implement fast dictionary searches [27] however it suffers from the limited accuracy of a single step (non-iterative) matched-filtering. KD-tree searches have been proposed to accelerate matched-filtering steps within an iterative reconstruction scheme [28]. However KD-trees are known to be

non-scalable and crucially dependent on a dimensionality-reduction preprocessing step and thus, using SVD subspace compression (as proposed) could introduce unfavourable accuracy vs. acceleration compromise in high-dimensional problems, as discussed above. Besides, the convergence results are heuristic due to the type of approximation used in [28]. We instead propose fast cover tree search iterations with an important feature of robustness against the curse-of-dimensionality [29, 21]. For low (intrinsic) dimensional manifold data cover trees have provable sub-linear search complexities and in addition we show that using such approximations within an iterative scheme can still result in monotone convergence (in general) and stable global reconstruction, under an embedding assumption.

Finally, we mention other schemes based on additional low rank priors [30, 13, 31, 32, 14], either through a fixed subspace model or costly singular values thresholding iterations. These methods are mostly validated on small problems and as discussed earlier a linear subspace i.e. low rank model will not scale to multi-parametric MRF setups with increased degree of non-linearities. Nonetheless and when usage is appropriate, CoverBLIP is also equipped with a (low rank) subspace model *option* based on a (pre-calculated) temporal factorization of the MRF dictionary. Also most recently deep learning approaches proposed to approximate match-filtering by compact neural networks during MRF reconstruction [33, 34, 35]. In our numerical comparisons we exclude these approaches and focus on purely dictionary-based reconstruction baselines.

3. MRF imaging model

MRF acquisitions follow a linear spatio-temporal model:

$$Y = P_\Omega F S(X) + \xi, \tag{1}$$

where $Y \in \mathbb{C}^{cm \times L}$ is the k-space measurements collected by c coils at $t = 1, \ldots, L$ temporal frames and corrupted by some noise ξ. The MRF image (to be recovered) is represented by a complex-valued matrix X of spatio-temporal resolution $n \times L$ i.e. n spatial voxels and L temporal frames[§]. The multi-coil sensitivity operator $S : \mathbb{C}^{n \times L} \to \mathbb{C}^{cn \times L}$ maps each temporal frame of X to c weighted copies according to the sensitivity maps of c head-coils used in a scanner. The sensitivity maps are identical for all temporal frames and are calculated off-line either through a separate calibration process or directly from the MRF measurements [36]. Throughout whenever we consider a single coil setup $c = 1$, we assume S to be an identity operator (i.e. $S(X) = X$) and thus the true sensitivities are absorbed by X. Moreover, F corresponds to a Fourier operator that maps spatial images (at each temporal frame and for each coil) to the corresponding k-space measurements. This operator might correspond to the FFT transform if a Cartesian grid is used for k-space sampling e.g. in [37, 16],

[§] The real and imaginary parts of X store net magnetizations across two transverse axes perpendicular to the static magnetic field.

or it might correspond to a Non-Uniform Fourier (NUFFT) transform [38] for non-Cartesian sampling patterns such as the variable density spirals used in [8, 10]. Finally, $P_\Omega : \mathbb{C}^{cn\times L} \to \mathbb{C}^{cm\times L}$ is the sub-sampling operator with respect to a set of *temporally-varying* patterns $\Omega = \bigcup_{t=1}^{L} \Omega_t$, where Ω_t stores $m < n$ k-space locations to be sampled at the time frame t. This pattern is identical for all coils at that given time frame. The linear system (1) is under-determined due to lack of sufficient measurements (i.e. $m < n$) which means without further assumptions it admits infinitely many solutions and therefore, in order to hope for a stable MRF reconstruction one needs to incorporate efficient and restrictive priors for this type of images.

3.1. Bloch dynamic model

The main source of measurements in Q-MRI are the per-voxel net magnetization of proton dipoles obtained from dynamic rotations of the external magnetic field induced by a radio frequency (RF) coil. These excitations are in the form of a sequence of Flip Angles (FA) $\{\alpha_t\}_{t=1}^{L}$ applied at certain time intervals known as the *repetition times* (TR) which could be a constant or varying across different time-frames $t = 1, \dots, L$. *Tissues with different NMR characteristics respond distinctively to these excitations.* A qualitative MRI approach studies the contrasts between different tissues in a single time frame. which is often dependent on the sequence type and the scanner. A Q-MRI approach rather fits a physical model to all spatio-temporal measurements and obtains the *absolute* NMR characteristics of the underlying tissues, however, at the cost of significantly longer acquisition times. Standard Q-MRI approaches e.g. [4, 5, 6, 7] use parameter-specific acquisition protocols that usually result in analytical time-trajectories such as $1 - 2\exp\left(-\frac{tTR}{T1}\right)$ or $\exp\left(-\frac{tTR}{T2}\right)$ to be fitted and recover the underlying parameters (here $T1$, $T2$ relaxation times).

The MRF framework relies on a similar principle, however, it adopts more complicated and sometimes random excitation patterns that are able to simultaneously encode different NMR parameters and produce more distinctive dynamic signatures in shorter acquisition times. The resulting temporal trajectories no longer follow simple analytic e.g. exponential forms and they require methods for approximating the solutions of the *Bloch differential equations* which capture the overall macroscopic dynamics of per-voxel magnetizations [9]. Denote by

$$\mathcal{B}(\Theta; TR, TE, \alpha) \in \mathbb{C}^L$$

the discrete-time Bloch response of a molecular structure with a set of intrinsic NMR parameters Θ to a specific excitation sequence of length L with a given FA pattern α, repetition TR and read-out TE times. The real and imaginary parts of $\mathcal{B}$ correspond to the amount of magnetizations across two transverse-plane components perpendicular to the external static magnetic field. For instance the IR-BSSFP sequence originally proposed for the MRF framework produces distinct magnetic responses for three parameters $\Theta = \{T1, T2, B0\}$ i.e. two relaxation times $T1$, $T2$ and the off-resonance

frequency B_0. Recent emerging MRF applications are designing sequences encoding a larger number of NMR characteristics such as $T2^*$, diffusion, perfusion and vascular properties (see e.g. [15, 16, 17, 18, 19, 20]).

Current MRF approaches discretize through a dense sampling the parameter space $\overline{\Theta} := [T1] \times [T2] \times [B0] \times \ldots$, simulate off-line the Bloch equations for all parameter combinations and generate a large dictionary of fingerprints $D = \{D_j\}_{j=1}^{d}$ where,

$$D_j := \mathcal{B}(\overline{\Theta}_j; TR, TE, \alpha) \quad \forall j = 1, \ldots, d, \tag{2}$$

and $d = \text{Card}(\overline{\Theta})$ is the total number of generated fingerprints (atoms). Under the *voxel purity* assumption each spatial voxel of the MRF image corresponds to a specific tissue with a unique NMR parameter and would approximately match to a temporal trajectory in the fingerprint dictionary.[‖] By incorporating a notion of signal intensity in this model the rows of the MRF image belong to a *cone* associated with the fingerprints (2). Denoting by X_v the v-th row of X i.e., a multi-dimensional spatial voxel, we have

$$X_v \in \text{cone}(D) \quad \forall v = 1, \ldots, n, \tag{3}$$

where the discrete cone of fingerprints is defined as follows:

$$\text{cone}(D) := \{x \in \mathbb{C}^L : x/\gamma \in D \quad \text{for some}\, \gamma > 0\}. \tag{4}$$

Here γ corresponds to the proton density which is generally non-uniformly distributed across spatial voxels.

3.2. Model-based MRF reconstruction

The discrete Bloch model in (3) plays a critical role in regularizing the inverse problem (1) and enabling stable MRF image reconstruction and parameter estimation. Following the model-based compressed sensing approaches such as [24, 23, 26], MRF reconstruction can be cast as minimizing the measurement discrepancy —through the forward model (1)—constrained by the per-voxel Bloch cone model:

$$\text{argmin}_X \sum_{t=1}^{L} \|Y_t - P_{\Omega_t} F S(X_t)\|_2^2 \quad s.t. \quad X_v \in \text{cone}(D) \quad \forall v = 1, \ldots, n.^{\dagger} \tag{5}$$

The recovered image sequence (solution) at each spatial voxel corresponds to a fingerprint representing uniquely the underlying NMR characterizations. As appeared in compressed sensing literature [40, 41, 42], it might be natural to think of incorporating

[‖] A number of works also consider mixture models for the MRF problem (see e.g. the supplementary part of [8] and a recent work [39]), however we keep the main focus of this paper is on imaging scenarios in which the signal can be approximated reasonably well through the pure voxel model.

[†] With a slight abuse of notation by $X_t \in \mathbb{C}^n$ we refer to the MRF image at its t-th temporal frame i.e. the t-th column of X, whereas by X_v we refer to the v-th row of X which is an L-dimensional spatial voxel. Also $Y_t \in \mathbb{C}^{cm}$ refers the k-space measurements collected at t-th repetition time.

additional priors to promote certain spatial regularities and/or low-rank structures (i.e. accounting for the correlations between neighbouring voxels or image patches) in order to improve reconstruction, see e.g. [12, 30, 28] in the MRF context. However care must be taken here, since solving a multi-constrained problem combined with the non-convex fingerprints cone (3) is often intractable and therefore despite possible empirical improvements —perhaps under good initializations—the results are likely to lack global convergence guarantees. In this paper we focus on problem (5) constrained by the cone of fingerprints.

A popular approach for solving compressed sensing problems is the Iterative Projected Gradient (IPG) algorithm [43, 44, 45]. IPG is a first-order algorithm suitable for big data applications and importantly it can also apply to globally solve problems with certain non-convex constraints [45, 24, 46]. Davies *et al.* [12] adopted this routine for the MRF reconstruction problem and named it Bloch Response Iterative Projection (BLIP). The BLIP algorithm iterates between gradient descent and (voxel-wise) model projection steps:

$$X^{k+1} = \mathcal{P}_{\mathcal{C}}\left(X^k - \mu_k \mathcal{A}^H\left(\mathcal{A}(X^k) - Y\right)\right), \tag{6}$$

where $\mathcal{A}(.) := P_\Omega F S(.)$ is the shorthand we use for the forward operator, $\mathcal{A}^H := S^H F^H P_\Omega^H(.)$ is the adjoint operator, $\{\mu_k\}$ is the sequence of step-sizes and $\mathcal{P}_{\mathcal{C}}(.)$ is the Euclidean projection operator onto the set $\mathcal{C}$ i.e.

$$\mathcal{P}_{\mathcal{C}}(x) \in \operatorname{argmin}_{x\in\mathcal{C}} \|x - u\|_2. \tag{7}$$

Note that throughout we use the shorthand $\|.\|$ to refer to the Euclidean norm i.e. the ℓ_2 norm of a vector or the Frobenius norm of a matrix. For the MRF problem and the constraint set $\mathcal{C}$ defined by (3) this projection is also called *matched-filtering*. After the gradient update $Z^k := X^k - \mu_k \mathcal{A}^H\left(\mathcal{A}(X^k) - Y\right)$, the matched-filtering step $X^{k+1} = \mathcal{P}_{\mathcal{C}}(Z^k)$ decouples into separate cone projections for each spatial voxel $v = 1, \dots n$ and is computed as follows:

$$j^* = \operatorname{argmin}_j \|Z_v - D_j/\|D_j\|\| \qquad \text{(nearest neighbour search)} \tag{8}$$

$$X_v^{k+1} = \mathcal{P}_{\text{cone}(D)}(Z_v) = \gamma_v D_{j^*} \qquad \text{(rescaling)} \tag{9}$$

where, $\gamma_v = \max\left(\text{real}(\langle Z_v, D_{j^*}\rangle)/\|D_{j^*}\|^2, 0\right)$ is the per-voxel proton density.

The non-iterative TM approach originally proposed in [8] corresponds to the first iteration of BLIP with zero initialization[‡]. However the iterative approach has shown to be more robust against shorter excitation sequences and acquisition times, where the atoms of the fingerprint dictionary become more coherent and difficult to be distinguished [12].

[‡] Throughout we assume zero initialization $X^{k=0} = \mathbf{0}$ for all iterative methods unless otherwise is specified.

3.3. Dimension-reduced subspace matched-filtering

Discretization of the multi-parameter space often results in very large size MRF dictionaries where the number of fingerprints d has an exponential relationship with the number of NMR characteristics and their quantization resolutions. Therefore, search strategies (e.g. brute-force) with linear complexity in d are a serious bottleneck to the exact matched-filtering steps at the heart of model-based approaches for solving (5). Current proposed solutions for the high dimensionality of the MRF problem rely on a (low rank) subspace compression step to reduce the matching computations [11, 28, 14]. Let $V \in \mathbb{C}^{L\times L}$ be the eigen-basis spanning the space of the fingerprint dictionary through the singular value decomposition (SVD) i.e. $\sum_{j=1}^{d} D_j(D_j)^H = V\Sigma V^H$, and $V_s \in \mathbb{C}^{L\times s}$ denotes the matrix of s-dominant eigenvectors. By assuming high (linear) correlations between fingerprints, there exists a reasonably small number $s \ll L$ for which one would have $D_j \approx V_s\widetilde{D}_j$ for all $j = 1, \dots d$, where $\widetilde{D}_j := V_s^H D_j \in \mathbb{C}^s$ and $\widetilde{D} := \{\widetilde{D}_j\}_{j=1}^{d}$ are the low-dimensional proxies for the original fingerprint dictionary. Assuming this, one can solve the following problem instead of (5) in lower dimensions:

$$\text{argmin}_{\widetilde{X}\in\mathbb{C}^{n\times s}} \sum_{t=1}^{L} \|Y_t - P_{\Omega_t} F S\big((\widetilde{X}V^H)_t\big)\|_2^2 \quad s.t. \quad \tilde{X}_v \in \text{cone}(\widetilde{D}) \quad \forall v = 1, \dots, n. \tag{10}$$

Note that if D is low-rank and fully spanned by V_s then $D = V_s\widetilde{D}$, $\text{cone}(D) = V_s\text{cone}(\widetilde{D})$ and by a change of variable we have $X = \widetilde{X}V^H$, and therefore both problems (5) and (10) become equivalent. Following the IPG routine for solving this problem, the gradient updates read

$$\widetilde{Z}^k = \widetilde{X}^k - \mu_k \mathcal{A}^H\Big(\mathcal{A}(\widetilde{X}^k V^H) - Y\Big)V, \tag{11}$$

where the matched-filtering $\widetilde{X}_v^{k+1} = \mathcal{P}_{\text{cone}(\widetilde{D})}(\widetilde{Z}_v^k)$ and the corresponding searches are performed in the compressed temporal domain, directly reducing the complexity of pairwise distance calculations. Such a compression scheme can also reduce the gradient step computations. One can write

$$\begin{aligned}\widetilde{Z}^k &:= \widetilde{X}^k - \mu_k \mathcal{A}^H\Big(\mathcal{A}(\widetilde{X}^k V^H)V\Big) + \mu_k \mathcal{A}^H(Y)V \\ &= \widetilde{X}^k - \mu_k S^H F^H\Big(P_\Omega^H P_\Omega\big(FS(\widetilde{X}^k)V^H\big)V\Big) + \mu_k \mathcal{A}^H(Y)V. \end{aligned}\tag{12}$$

The last line follows from expanding $\mathcal{A}$ and it holds since both the multi-slice Fourier transform F and the coil sensitivity operator S act identically across all time-frames and thus they commute with the temporal compression operators V, V^H. As a result, the main computations for conducting the gradient updates (12) i.e., the middle term, comes from the forward-backward Fourier operations across a smaller number $s < L$ of (compressed) temporal frames plus the cost of applying compression-decompression operations V, V^H. Depending on how well a low rank model can approximate the dictionary i.e. how small would s be, the overall gradient computations can drop by

using such subspace compression, particularly when F corresponds to expensive NUFFT transforms in non-Cartesian acquisition schemes. We empirically observe that V, V^H operations would not bring a major overhead in total computations.

The idea of using subspace compressions has been applied to accelerate brute-force searches in the single-stage TM method [11] as well as an iterative algorithm [28] to boost the performance of fast but non-scalable searches based on KD-trees. The later approach is totally reliant on such a compression pre-processing since it is well understood that KD-trees are inefficient in high-dimensional (ambient) search spaces. Beside these advantages, we would like to remind the reader about our discussion in sections 2 (see also our numerical experiments in Section 5), that methodologies purely relying on subspace dimensionality-reduction are prone to an unfavourable compromise in their estimation accuracies when applied to multi-parametric MRF dictionaries with increased non-linear complexities and growth in data population.

4. Accelerated MRF reconstruction with scalable tree searches

Accelerating the Nearest Neighbour Search (NNS) is a fundamental problem in computer science and it has a long historical literature. Successful proposed approaches are based on building tree structures which hierarchically partition large datasets and then use branch-and-bound algorithms for fast NNS (see e.g. [47, 48, 49, 21]). KD-trees —which are the multi-dimensional generalization of binary searches —are perhaps the most widely-known classical structure for fast searches [47]. They consist of partitioning datasets across ambient coordinate axes and therefore do not efficiently adapt to complicated low-dimensional structures of datasets embedded into high (ambient) dimensions. A dimensionality reduction step is inevitably necessary when using KD-trees since they are non-scalable and their search complexity rapidly grows in high-dimensional problems [50]. Modern search algorithms circumvent the curse-of-dimensionality by using i) tree structures that could efficiently benefit from the low *intrinsic* dimensionality of natural datasets, which is a key assumption in machine learning, and ii) low-complexity algorithms for performing the search *approximately* i.e. Approximate Nearest Neighbour Search (ANNS).

4.1. Cover trees

We are particularly interested in deploying a recent data structure known as *Cover tree* [21]. A cover tree is a levelled tree which forms covering nets for data at multiple (dyadic) resolutions. Several key growth properties such as cover tree's depth, width, and importantly the overall search complexity are characterized by the intrinsic dimension of data [29]. There is a great motivation behind using cover trees searches for the MRF reconstruction problem. The manifold $\mathcal{M} := \mathcal{B}(\Theta) \in \mathbb{C}^L$ corresponding to the solutions of Bloch equations is parametrized by a small number $\text{Card}(\Theta) \ll L$ of parameters; an observation which implies the resulting fingerprint dictionary will

have a low-dimensional intrinsic structure. In our supplementary materials we highlight in further details key properties that make this structure ideal for accelerated and scalable searches within iterative MRF reconstruction. To summarize, for datasets with low intrinsic dimensions cover trees have appealing construction-time complexities and memory requirements and notably, they can achieve a logarithmic search complexity in terms of data population without needing an explicit a-priori knowledge of the data structure nor a dimensionality reduction preprocessing.

Definition 1. *Given a dataset D, a query point p (which might not belong to D) and $\varepsilon \geqslant 0$, then a point $q \in D$ from dataset is a $(1+\varepsilon)$-approximate nearest neighbour of p if it holds:*

$$\|p - q\| \leqslant (1+\varepsilon) \min_{u \in D} \|p - u\|. \tag{13}$$

Authors in [21, 51] proposed a branch-and-bound procedure for $(1+\varepsilon)$-ANNS using a cover tree structure $\mathcal{T}(D)$ on data. This algorithm (also detailed in our supplementary materials) can be initialized with a current estimate $q_c \in D$

$$q \leftarrow (1+\varepsilon)\textbf{-ANNS}\left(p, \mathcal{T}(D), q_c\right), \tag{14}$$

where the search output jointly satisfies (13) and $\|q - p\| \leqslant \|q_c - p\|$. Depending on the initialization quality i.e. q_c's distance to approximate solutions, this routine can enjoy further accelerations due to ruling out many branches at top levels of the tree. Also, one should distinguish between the exact tree NNS (case $\varepsilon = 0$) and performing a brute-force search. Although they both perform an exact NNS, the complexity of 14 is empirically shown to be way less in practical datasets.

4.2. CoverBLIP algorithm

Approximation plays a key role in accelerating the nearest neighbour searches and breaking the curse-of-dimensionality [50]. Motivated by the low-dimensional (manifold) structures present in the MRF dictionary, we propose to accelerate iterative matched-filtering steps within the BLIP algorithm by using cover tree's $(1+\varepsilon)$-ANNS approximate searches. Algorithm 1 outlines the proposed Cover tree BLoch Iterative Projection (CoverBLIP) procedure for accelerated MRF reconstruction. We replace the exact NNS step (8) in the cone projection with the following approximation, $\forall v = 1, \ldots, n$:

$$D_{j^*_{k+1}} = (1+\varepsilon)\textbf{-ANNS}\left(Z_v/\|Z_v\|, \mathcal{T}(D), D_{j^*_k}\right), \tag{15}$$

which uses search algorithm 14 for a given inaccuracy level $\varepsilon \geqslant 0$. We denote by $\mathcal{T}(D)$ the cover tree structure built for the normalized fingerprint dictionary. At each iteration CoverBLIP uses previously selected fingerprints (i.e. $D_{j^*_k} = X^k_v/\gamma_v$ for each voxel) to initialize the ANNS searches. This has two positive impacts: i) the search achieves further acceleration especially, close to the converging point of the algorithm, because with an initialization close to the ANNS solution the branch-and-bound procedure can effectively rule out many branches at higher levels of the tree and thus keep the

Algorithm 1 CoverBLIP$(Y, \mathcal{T}(D), \mathcal{A}, \mu)$

Inputs: k-space measurements Y, cover tree structure $\mathcal{T}(D)$ constructed for the normalized fingerprint dictionary D, forward operator $\mathcal{A} := P_\Omega FS$ and its corresponding adjoint operator $\mathcal{A}^H$, initial step-size μ.
Initialization: $k = 0$, $X^0 = \mathbf{0}$, $\mu_k = \mu \ \forall k = 1, 2, \ldots$
while stopping criterion = false **do**
  $Z = X^k - \mu_k \mathcal{A}^H \left(\mathcal{A}(X^k) - Y\right)$ #(gradient update)
  for $v = 1, \ldots, n$ **do** #(per-voxel approximate model projection)
    $D_{j^*_{k+1},v} = (1+\varepsilon)$**-ANNS** $\left(Z_v/\|Z_v\|, \mathcal{T}(D), D_{j^*_k,v}\right)$ #(cover tree's ANNS)
    $\gamma_v = \max\left(\text{real}(\langle Z_v, D_{j^*_{k+1},v}\rangle)/\|D_{j^*_{k+1},v}\|^2, 0\right)$
    $X^{k+1}_v = \gamma_v D_{j^*_{k+1},v}$ #(rescaling)
  if $\mu_k \geqslant \frac{\|X^{k+1}-X^k\|^2}{\|\mathcal{A}(X^{k+1}-X^k)\|^2}$ **then** #(adaptive step-size shrinkage)
    $\mu_k = \mu_k/2$
  else
    $k = k + 1$
$\overline{\Theta}_v \leftarrow$ look-up-table$\left(D_{j^*_{k+1},v}\right), \forall v$
return reconstructed MRF image X^{k+1}, parameter maps $\overline{\Theta}$, proton density γ

candidates set very small, and ii) the $(1+\varepsilon)$-ANNS algorithm would produce non-expansive outputs i.e. $\forall v$ we have

$$\|Z_v/\|Z_v\| - D_{j^*_{k+1}}\| \leqslant \|Z_v/\|Z_v\| - D_{j^*_k}\|, \tag{16}$$

which as will be discussed in the next part it is a key property to guarantee the monotone convergence of CoverBLIP. Note that we feed the search algorithm with the normalized gradient updates $Z_v/\|Z_v\|$. Since dictionary atoms are normalized the search outcome is invariant with respect to the query rescaling, however from the complexity perspective one would gain in computation time by searching a query within a closer range to datasets' hypersphere. We also observed in our experiments that this trick leads to better accelerations.

Convergence is tied to a proper choice of the step-size sequence. We follow the adaptive scheme proposed in [52] which starts from a large initial step size and shrinks this choice by a division factor $\zeta > 1$ e.g., half of the previous step size by setting $\zeta = 2$, until meeting the following criteria at each iteration k:

$$\mu_k < \frac{\|X^{k+1} - X^k\|^2}{\|\mathcal{A}(X^{k+1} - X^k)\|^2} \tag{17}$$

where again, $\mathcal{A}(.) := P_\Omega FS(.)$ is the shorthand for the forward operator. This condition is another important ingredient to guarantee the convergence of CoverBLIP iterations, which supported by some extra assumptions will also imply a robust reconstruction i.e. near global convergence. We will discuss this point in further details in the next

section. After the first iteration we can also use the following energy ratio between measurements and our first estimation i.e. $\kappa = \|Y\|/\|\mathcal{A}(X^1)\|$ in order to rescale the first iteration $X^1 \leftarrow \kappa X^1$ and set an appropriate range (e.g. large enough) for the initial step size $\mu \leftarrow \kappa\mu$.

When applicable —and with a possible compromise in the accuracy —a temporal subspace compression similar as explained in Section 3.3 can be optionally included to further shrink dimensions of Z_v, X_v, D_j across the dominant SVD components $V_s \in \mathbb{C}^{L\times s}$ of the MRF dictionary. In this case one has to build a cover tree structure for the normalized dimension-reduced dictionary $\widetilde{D}$, update the gradient step in Algorithm 1 by the expression (11), and for the step-size expression (17) would change to

$$\mu < \frac{\|\widetilde{X}^{k+1} - \widetilde{X}^k\|^2}{\|\mathcal{A}(\widetilde{X}^{k+1}V_s^H - \widetilde{X}^k V_s^H)\|^2}. \tag{18}$$

The updated gradient step might also introduce a compromise between cheaper distance evaluations during the search steps (i.e. in $\mathbb{C}^s$ rather than $\mathbb{C}^L$) and a computation overhead due to applying iteratively compression and decompression, as previously highlighted in Section 3.3.

The approximate projection step presented in Algorithm 1 (i.e. lines 6 and 7) assumes that proton densities are real and positive valued quantities. A phase-alignment heuristic similar to [28] can be used to extend this framework to complex-valued proton densities. This approach approximates dictionary atoms with fingerprints having constant complex phases across temporal domain. Complex angles corresponding to the first principal component i.e. $\widetilde{D}_{s=1} = V_{s=1}^H D$, are then used to align dictionary atoms. Similarly, at each iteration in line 6 we align phases of the gradient update used for the search step; In our experiments we use the complex angles of the dominant compressed image i.e. angle($\widetilde{X}_{s=1}$) for temporal phase-alignment. Empirical results applying this approach are demonstrated for our volunteer data experiments in Section 5.2.

4.3. Convergence of CoverBLIP

The analysis in this part covers the behaviour of a wide class of *inexact* IPG algorithms for solving linear inverse problems where the forward operator $\mathcal{A}$ and the set $\mathcal{C}$ of signal model could be regarded in general forms and not necessary customized for the MRF recovery problem.

A previous work [26] studied the stability of the inexact IPG algorithms with respect to several forms of approximations on gradient and projection updates. Here we focus on iterative algorithms that use the following notion of relative approximate projection step i.e. for an $\varepsilon \geqslant 0$ we define

$$\widetilde{\mathcal{P}}_{\mathcal{C}}^{\varepsilon}(x) \in \left\{ u \in \mathcal{C} : \|u - x\| \leqslant (1+\varepsilon) \inf_{u' \in \mathcal{C}} \|u' - x\| \right\}. \tag{19}$$

Example 1. Following Definition 1, the $(1+\varepsilon)$-ANNS search algorithm is an approximate

projection of type (19) onto a discrete set of points $\mathcal{C} := D$ in a dataset e.g. a signal model which is used for data-driven inverse problems [26].

Example 2. Notably for projection onto $\mathcal{C} := \text{cone}(D)$, if we replace the exact search step in (8) with an approximate $(1+\varepsilon)$-ANNS search, we obtain an approximate cone projection $\mathcal{P}^{\varepsilon}_{\text{cone}(D)}(.)$ satisfying definition (19). Steps 6 to 8 in CoverBLIP Algorithm 1 are indeed implementing such an approximate projection onto the cone associated with the MR fingerprints using fast cover tree searches.

The corresponding inexact IPG iterations, including the CoverBLIP algorithm as a particular case, are as follows:

$$X^{k+1} = \mathcal{P}^{\varepsilon}_{\mathcal{C}}\left(X^k - \mu_k \mathcal{A}^H\left(\mathcal{A}(X^k) - Y\right)\right), \tag{20}$$

We now follow this section by discussing two types of guarantees for the inexact IPG. The first type makes an embedding assumption on $(\mathcal{A}, \mathcal{C})$ and provides a robust signal recovery result which in turn implies an interesting near global convergence guarantee for arbitrary signal models $\mathcal{C}$ including the non-convex conic constraints in the MRF problem. The second form of our analysis does not make an embedding assumption and only relies on an adaptive step-size scheme to ensure criteria (17) holds and guarantees local convergence of the inexact IPG algorithm.

The following embedding assumption plays a critical role in our stable signal recovery result [23, 25, 26]:

Definition 2. *A forward operator $\mathcal{A}$ is bi-Lipschitz with respect to a set $\mathcal{C}$, if $\forall x, x' \in \mathcal{C}$ there exists constants $0 < \alpha \leqslant \beta$ such that*

$$\alpha\|x - x'\|^2 \leqslant \|\mathcal{A}(x - x')\|^2 \leqslant \beta\|x - x'\|^2. \tag{21}$$

Equipped with this notion the following result states that when we have a good measurement consistency i.e. when $\min_{X\in\mathcal{C}} \|Y - \mathcal{A}(X)\|$ is small, then a near global convergence could be achieved using inexact iterations [26, 1]:

Theorem 1. *Assume $(\mathcal{A}, \mathcal{C})$ is bi-Lipschitz and that for a given $\varepsilon \geqslant 0$ and some constant $\delta \in [0, 1)$ it holds*

$$\sqrt{\varepsilon + \varepsilon^2} \leqslant \delta\sqrt{\alpha}/|||\mathcal{A}||| \quad \textit{and} \quad \beta < (2 - 2\delta + \delta^2)\alpha,$$

where $|||\mathcal{A}|||$ denotes the spectral norm of $\mathcal{A}$. Set the step size $\mu_k = \mu, \forall k$ such that

$$\left((2 - 2\delta + \delta^2)\alpha\right)^{-1} < \mu \leqslant \beta^{-1}.$$

The sequence generated by Algorithm (20) *obeys the following bound:*

$$\|X^k - X_0\| \leqslant \rho^k \|X_0\| + \frac{\kappa_w}{1 - \rho} w \tag{22}$$

where $X_0 = \text{argmin}_{X\in\mathcal{C}} \|Y - \mathcal{A}(X)\|$, $w = \|Y - \mathcal{A}(X_0)\|$ and

$$\rho = \sqrt{\frac{1}{\mu\alpha} - 1} + \delta, \quad \kappa_w = 2\frac{\sqrt{\beta}}{\alpha} + \sqrt{\mu}\delta.$$

Remark 1. Theorem 1 guarantees a linear convergence behaviour for inexact iterations. As a result after a finite $K = O(\log(\tau^{-1}))$ number of iterations Algorithm (20) achieves the solution accuracy $\|X^K - X_0\| = O(w) + \tau$ for any $\tau > 0$.

Remark 2. Under a properly conditioned bi-Lipschitz embedding as assumed in Theorem 1 the inexact algorithm achieves a solution accuracy comparable to that of the exact IPG algorithm. By increasing $\varepsilon > 0$ we require better embedding conditions as compared to the exact iterations (i.e. the case $\varepsilon, \delta = 0$). Although, increasing ε slows down the rate ρ of linear convergence, it could facilitate significantly cheaper computations per iteration. In other words, approximation trades-off against the embedding conditions, rate of convergence and computation time, but not against the order of the solution accuracy.

The following proposition (see the proof in the supplementary materials) says that by using the adaptive shrinkage scheme described in the previous part we can find a good step size in a finite (logarithmic) number of sub-iterations:

Proposition 1. *Following the iterative step-size shrinkage scheme with the initial size μ and division factor $\zeta > 1$, the chosen step size $\mu_k, \forall k$ meets the criteria* (17) *and satisfies the following bound:*

$$(\zeta\beta)^{-1} \leqslant \mu_k \leqslant \alpha^{-1}. \tag{23}$$

in a finite number $\lceil \log_\zeta(\beta\mu) \rceil + 1$ of iterations.

The following theorem establishes a stable reconstruction guarantee (i.e. near global convergence) for the inexact IPG algorithm by using the adaptive step size shrinkage scheme:

Theorem 2. *Assume $(\mathcal{A}, \mathcal{C})$ is bi-Lipschitz, and that for given $\varepsilon \geqslant 0$, $\zeta > 1$ and some constant $\delta \in [0, 1)$ it holds*

$$\sqrt{\varepsilon + \varepsilon^2} \leqslant \delta\sqrt{\alpha}/|||\mathcal{A}||| \quad \textit{and} \quad \zeta\beta < (2 - 2\delta + \delta^2)\alpha.$$

Following the adaptive step-size scheme with shrinkage factor ζ, the sequence generated by Algorithm (20) *obeys the error bound* (22) *where,*

$$\rho = \sqrt{\frac{\zeta\beta}{\alpha} - 1} + \delta, \quad \kappa_w = 2\frac{\sqrt{\beta}}{\alpha} + \frac{\delta}{\sqrt{\alpha}}.$$

The proof architecture is similar to the proof of [26, Theorem 2], however, this result does not a priori assume $\mu_k \leqslant 1/\beta$ as there or in Theorem 1 of this paper. For the sake of completeness we provide detailed proof of Theorem 2 in the supplementary materials.

Remark 3. Without an a-priori knowledge of the embedding constants, the inexact IPG algorithm with adaptive step-size exhibits a similar linear convergence behaviour towards the global minima as in Theorem 1. The closer ζ is chosen to one, the embedding condition and the rate of convergence become more comparable to Theorem 1, however at the increased cost of more shrinkage sub-iterations.

Remark 4. Theorem 1 generalizes results in [52] in two ways: i) the set $\mathcal{C}$ of constraints are general and not restricted to sparse signals, and ii) results here establish robustness against inexact projection updates. Notably when no approximation is used $\varepsilon, \delta = 0$, Theorem 1 relaxes the embedding conditions in [52, Theorem 3] which required $\zeta\beta < 8/7\alpha$.

Finally, we consider a general convergence result which holds even in the absence of the bi-Lipschitz embedding assumption. We additionally assume that the approximate projection produces non-expansive updates with respect to the previous iterations i.e. $\forall k$ and gradient updates $Z^k := X^k - \mu_k \mathcal{A}^H \left(\mathcal{A}(X^k) - Y\right)$ it holds :

$$\|\mathcal{P}_{\mathcal{C}}^{\varepsilon}(Z^k) - Z^k\| \leqslant \|X^k - Z^k\|. \tag{24}$$

Example 3. The (1+ε)-ANNS update (15) in CoverBLIP and the associated approximate cone projection satisfy the non-expansiveness property (24), thanks to initializing the search algorithm with previous iteration.[§]

The following result (proof in the supplementary materials) guarantees monotone convergence of Algorithm (20) since the cost function $\|Y - \mathcal{A}(X)\| \geqslant 0$ is lower bounded:

Theorem 3. *Assume the approximate projections are non-expansive and the step-size satisfies* (17)*. Algorithm* (20) *produces a non-increasing and convergent sequence* $\|Y - \mathcal{A}(X^k)\|$.

Note that determining the bi-Lipschitz conditioning i.e. constants α and β and hence an admissible interval for choosing the step-size (as suggested in Theorem 1) is a combinatorial problem in general. For a certain class of *random* sampling schemes used in compressed sensing theory however it is possible to derive those constants with high probability, see e.g. [53]. Applied to the MRF problem, it has been shown in [12, Theorem 1] that if sampling patterns Ω_t sub-select uniformly at random large enough number of rows (or columns) of the k-space —a sampling protocol referred to as the random Echo Planar Imaging (EPI) —then the resulting forward model $\mathcal{A}$ is bi-Lipschitz, and a fixed choice of step-size equal to the compression factor

$$\mu_k = n/m, \quad \forall k$$

guarantees stable reconstruction. Randomized acquisition schemes are however not currently popular in practical MRF setups, leading to pronounce more the importance of theorems 2 and 3. In Theorem 2 one does not need to explicitly obtain the bi-Lipschitz constants, however if the forward model happens to satisfy a proper embedding condition then the adaptive step-size scheme is determined to make a proper choice (within the corresponding admissible interval) which guarantees near global convergence. Otherwise in the absence of any embedding assumption Theorem 3 ensures monotone convergence of the non-convex projected iterations i.e. the stability of the algorithm.

[§] In general one could easily incorporate property (24) by the following update in Algorithm (20): $X^{k+1} = \text{argmin}_{u \in \{\mathcal{P}_{\mathcal{C}}^{\varepsilon}(Z^k), X^k\}} \|u - Z^k\|$.

Algorithm	Description
BLIP	Iterative reconstruction (6) using exact brute-force searches for the matched-filtering [12]
Template Matching (TM)	Non iterative matched-filtering reconstruction using exact brute-force searches [8] (i.e. the first iteration of BLIP)
KDBLIP	Iterative reconstruction similar to [28] using KD-tree's ANNS for the matched filtering (Approximation level is controlled by the number of *checks* which specifies the maximum leafs to visit during the search. Higher checks values give better search precision, but also take more time)
CoverBLIP	Iterative reconstruction Algorithm 1 using cover tree's $(1+\varepsilon)$-ANNS for the matched filtering (Approximation level is controlled by $\varepsilon \geqslant 0$ which bounds the search precision according to Definition 1. Smaller ε would give better search precision, but also take more time).

Table 1: Algorithms used for validations and comparisons.

5. Numerical experiments

In this section we evaluate and compare the performance of the proposed CoverBLIP algorithm against dictionary-based MRF reconstruction baselines listed in Table 1. Experiments are conducted using MATLAB on a moderate desktop with 8 CPU-Cores and 32 GB RAM. For BLIP and TM algorithms the exact NNS is calculated using MATLAB's matrix product. KDBLIP iterations use randomized KD-tree searches implemented by the FLANN package [54]. Our CoverBLIP algorithm uses a parallel MATLAB interface to an existing implementation of the cover tree's $(1+\varepsilon)$-ANNS in [55].[‖] We do not believe this implementation is as optimized as that of the FLANN package for KD-tree searches and thus any reconstruction time comparisons (if not unfair) must take this point into account.

a) Temporal subspace compression option: As discussed in Section 3.3 all considered methods here can use a temporal compression option where the corresponding subspaces are the $s \leqslant L$ dominant SVD components of the fingerprint dictionary. This option has the advantage of reconstructing smaller objects i.e. MRF images, accelerated gradient updates i.e. forward and adjoint Fourier operations, and performing searches in low dimensional (ambient) space. The later is particularly beneficial for non-scalable search schemes such as KD-trees.

b) Datasets: Two sets of experiments are conducted: one using the synthetic Brainweb digital phantom with available Ground Truth (GT) maps [56], and the other using *in-vivo* scans of a healthy volunteer's brain which appeared in the original work of Ma *et al.* [8]. Both experiments use the Inversion Recovery (IR) Balanced SSFP acquisition sequence of length $L = 1000$, however with different flip angles and repetition times TR. The resulting temporal signals from the IR-BSSFP sequence encode three NMR parameters $\Theta = \{T1, T2, B0\}$ i.e. the relaxation times and the off-resonance frequency. For each experiment and given FA and TR patterns a fingerprint dictionary is created as

[‖] Implementations related to this work are available online at: `http://github.com/mgolbabaee/CoverBLIP`.

in [8] by solving discrete-time Bloch equations for combinations of the NMR parameters.

c) Evaluation metrics: The normalized solution MSE (NMSE) is measured as $\frac{\|\hat{X}-X_0\|}{\|X_0\|}$ where $X_0, \hat{X}$ are the ground truth and the reconstructed MRF images, respectively. The NMR parameter estimation accuracies are measured e.g. for the $T1$ case, as follows:

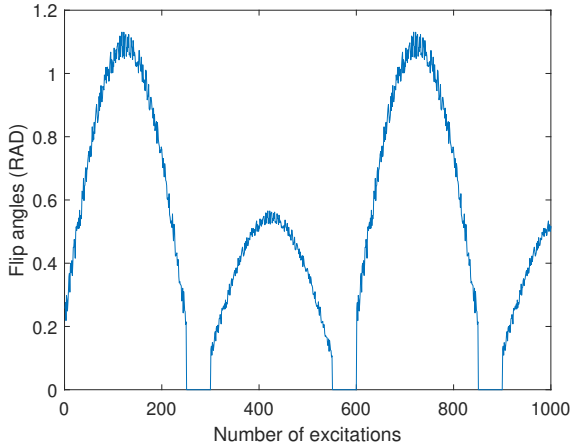
$$T1 \text{ accuracy} := 1 - \frac{1}{\text{Card}(\mathcal{N})} \Sigma_{v \in \mathcal{N}} \frac{|\hat{T1}(v) - T1(v)|}{T1(v)}$$

where v is the number of voxels within a masked region $\mathcal{N}$ defining the object of interest. The mask is obtained by contouring the output proton density (PD) map of the brain and removing empty voxels where the quantitative values are undefined. $T1(v)$ represents the ground truth $T1$ value for the v-th voxel and $\hat{T1}(v)$ is the corresponding estimated value.

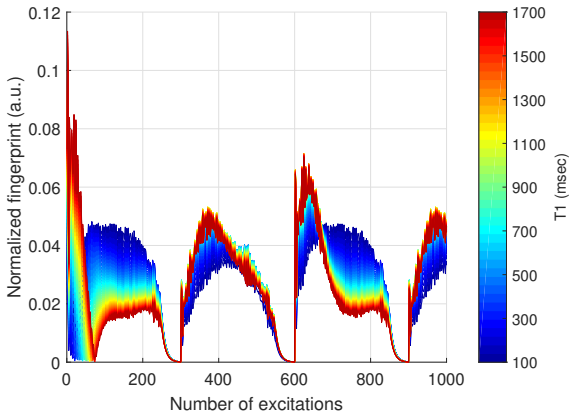
d) Computational cost: To have a fair comparison between computational complexities of the considered methods —and independent from how optimally they are implemented —we measure *total search costs* in addition to the reconstruction times. The cost measures the total number of computed pairwise distances (multiplied by the search dimension i.e. either L or $s \leqslant L$ when using subspace compression) for performing the NNS or ANNS steps within (iterative) matched-filtering until the algorithm converges. For all iterative methods the maximum number of iterations is set to 50 and the algorithm is stopped earlier if the relative progress in minimizing the objective function of (5) (or (10) when using subspace compression) is less than 10^{-6}.

5.1. Brainweb phantom with multi-shot EPI acquisition

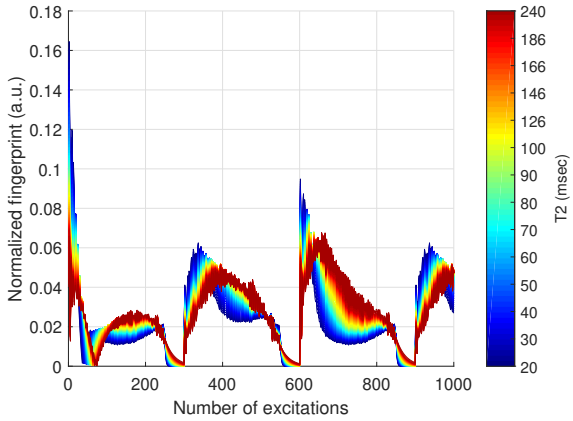
In this part we compare the performance of different MRF recovery methods on a synthetic dataset X_0 of size $(n = 256^2) \times (L = 1000)$ generated from the numerical Brainweb phantom (see details in supplementary materials). Figure 1 illustrates the FA pattern used for synthesizing this dataset and the dictionary of $d = 314160$ fingerprints used for reconstruction in this experiment. For k-space sampling we simulate a similar protocol to the recently proposed multi-shot Echo Planar Imaging (EPI) for MRF acquisition [37]. This protocol is based on a Cartesian grid Fourier sub-sampling where at each repetition time 16 out of 256 lines (with uniform spacing) from the k-space are simultaneously measured. In the next time frame the sixteen-shot sampling pattern Ω_t will be shifted by one line and so on. As a result we are dealing with reconstructing a 16x-fold undersampled data. We consider a single coil acquisition $S(X) = X$ and white Gaussian noise of 50 dB SNR added to the k-space measurements. Figure 2 illustrates the ground truth MRF images X_0 and the highly aliased back-projected images (BPI) i.e. $X' = \mathcal{A}^H(Y)$ using this sampling protocol.



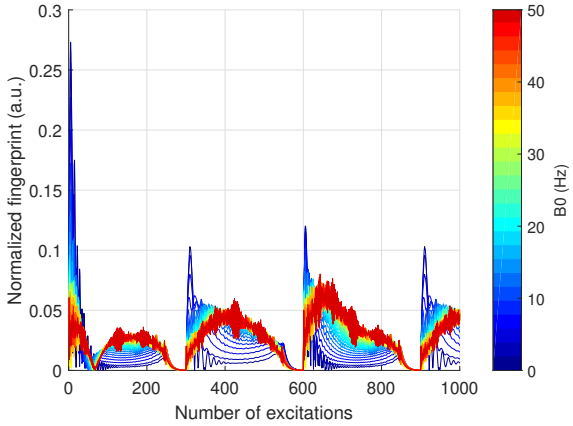
(a) Flip angles (FA)



(b) $T2 = 100$ (msec), $B0 = 20$ (Hz)



(c) $T1 = 1100$ (msec), $B0 = 20$ (Hz)



(d) $T1 = 1100$ (msec), $T2 = 100$ (msec)

Figure 1: The IR-BSSFP dictionary generated from a set of pseudo-random FAs shown in Figure (a) and fixed $TR = 10$ (msec). This dictionary encodes $T1$ and $T2$ relaxation times and off-resonance frequency $B0$. Figures (b)-(d) show the magnitude of the complex fingerprints (i.e. dictionary atoms) for different parameter combinations.

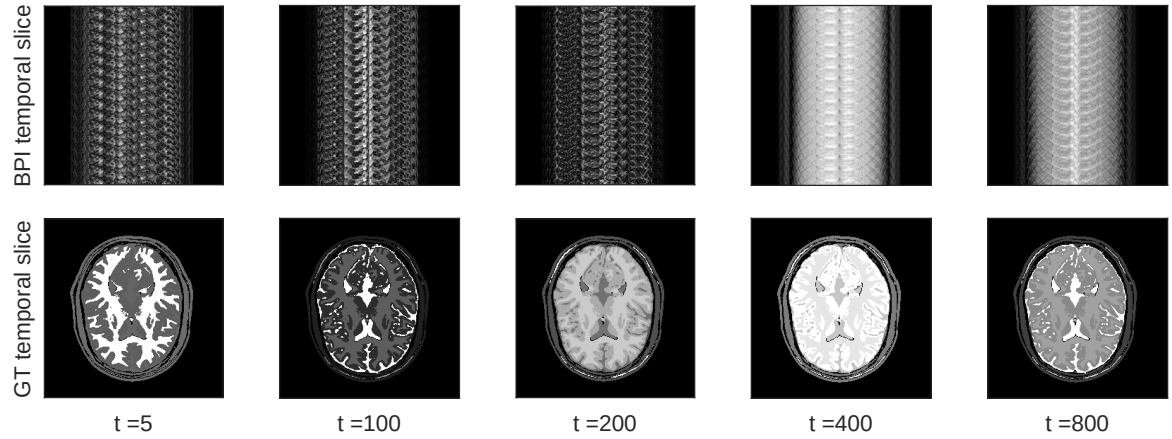


Figure 2: Ground truth MRF images generated from the Brainweb phantom (bottom row) and the highly aliased BPIs (top row) $X' = \mathcal{A}^H(Y)$ across different time frames.

Algorithm	checks/ε	NMSE	T1 acc. (%)	T2 acc. (%)	df acc. (%)	search cost	search time (sec)	iter.	total runtime (sec)
Temporal compression $s = 20$									
BLIP	-	1.563e-1	94.2	85.3	75.3	1.11e13	1.08e3	14	1.11e3
CoverBLIP	0.4	1.295e-1	94.6	89.2	81.2	3.94e9	2.74e1	14	5.65e1
KDBLIP	256	1.355-1	94.2	85.8	79.6	5.83e9	6.62e1	14	9.34e1
No temporal compression									
BLIP	-	5.327e-3	99.4	98.5	84.3	6.59e14	1.33e4	20	1.34e4
CoverBLIP	0.4	5.300e-3	99.4	98.5	84.3	4.86e11	3.45e2	25	4.93e2
KDBLIP	256	1.727e-1	92.7	78.7	67.4	3.05e12	4.49e2	38	6.96e2

Table 2: Comparisons between iterative methods with/without using subspace compression in terms of reconstruction NMSE, parameter estimation accuracy, search cost/time, number of iterations and total reconstruction time.

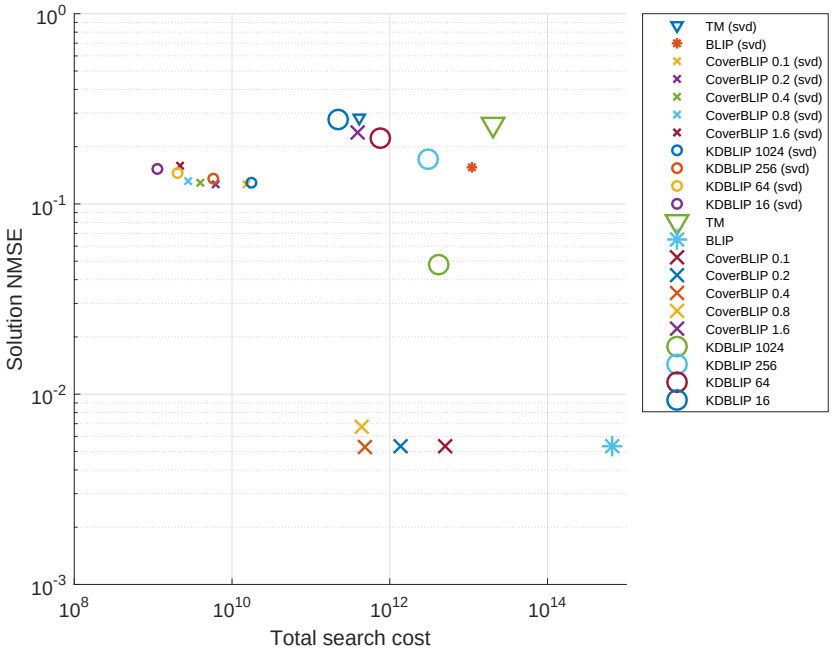
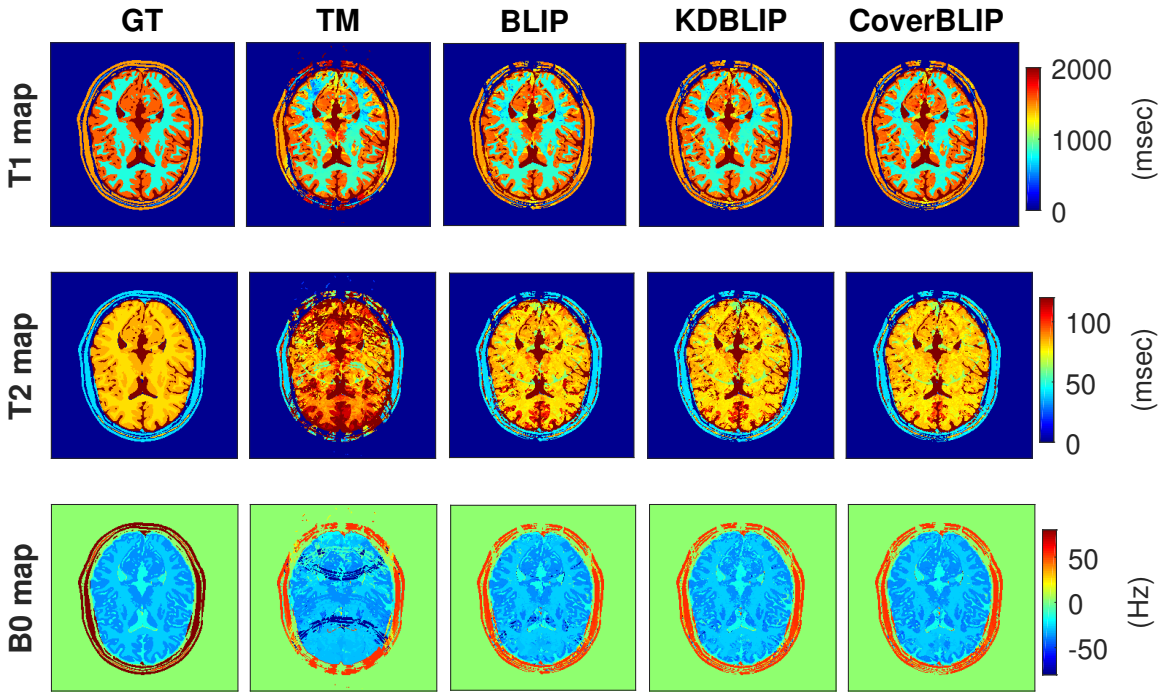
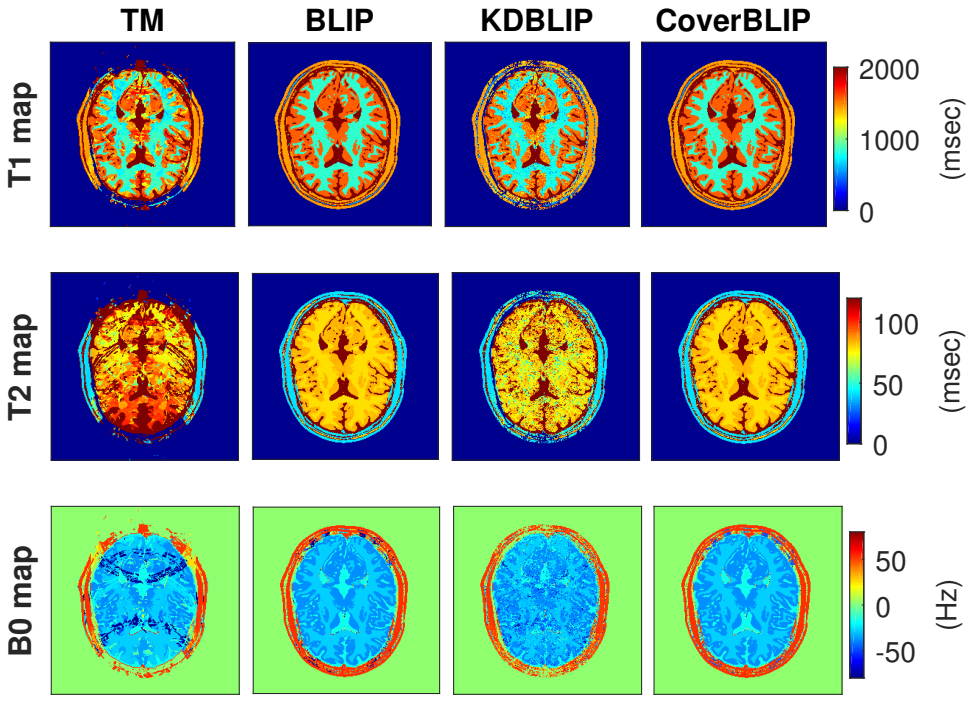


Figure 3: Total search cost vs. solution accuracy of the non-iterative TM and iterative BLIP algorithms using brute-force searches, and inexact iterative algorithms CoverBLIP and KDBLIP using fast tree searches. Two scenarios of applying temporal SVD compression where $s = 20$, and using no temporal compression are compared.

5.1.1. Results We report reconstruction times, total search (projection) costs, image reconstruction errors and parameter estimation accuracies in Figure 3 and Table 2. We also show the reconstructed parameter maps in Figure 4. For the KDBLIP algorithm we vary the KD-tree's search accuracy level by choosing checks $= \{1024, 256, 64, 16\}$. For the CoverBLIP algorithm we also test different $(1+\varepsilon)$-ANNS search approximations by



(a) Temporal compression ($s = 20$)



(b) No temporal compression

Figure 4: The ground truth (GT) and reconstructed $T1$, $T2$ and $B0$ maps for the numerical Brainweb phantom acquired by the (simulated) multi-shot EPI protocol with x16-fold under-sampling. KDBLIP and CoverBLIP iterations use search accuracies checks $= 256$ and $\varepsilon = 0.4$, respectively. Figures (a) and (b) compare the estimated maps with/without using the SVD based temporal compression for all tested algorithms.

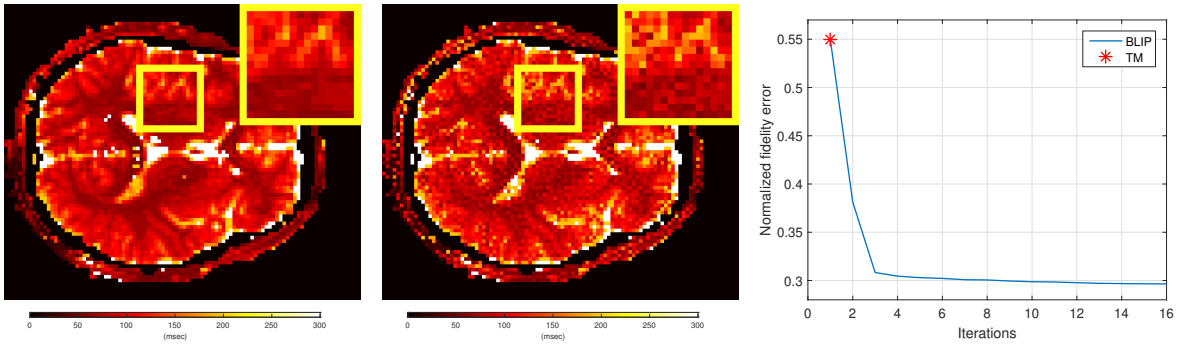
choosing $\varepsilon = \{0.1, 0.2, 0.4, 0.8, 1.6\}$. We initialize the step-size of the iterative schemes by the compression factor $\mu = n/m$ which empirically turns out to satisfy criteria (17) in most iterations and requires one shrinkage sub-iteration for the rest (see discussions in Section 4.3).

Temporal SVD compression ($s = 20$) accelerates the runtimes of all tested methods within 1-2 orders of magnitude (Table 2), however such an aggressive compression leads to poor parameter reconstructions (Figure 4(a)). Focusing on the non-compressed regime, we can see that Template Matching (TM) cannot achieve a good accuracy compared to the iterative methods (Figures 3 and 4(b)). The BLIP algorithm addresses this issue however at a high computational cost of iterating exact brute-force searches. Note that since the multi-shot EPI acquisition uses a Cartesian sampling, F in the forward model (5) corresponds to a FFT operator with fast gradient updates. As a result, and as can be observed in Table 2, projections (i.e. searches) dominate the runtimes of the iterative methods and thus accelerating this step would directly improve the total reconstruction time. CoverBLIP does so by using inexact cover tree searches (e.g. $\varepsilon = 0.4$) and achieves the best reconstruction time-accuracy (also search cost-accuracy) in all cases. Remarkably, CoverBLIP reports a similar accuracy to BLIP iterations however with 3 orders of magnitude less search cost and 27x-fold acceleration in the reconstruction time. Notably, the total cost of CoverBLIP inexact searches does not exceed that of a single stage brute-force search in TM (Figure 3).

When using temporal compression—a favourable case for the KD-tree searches—KDBLIP with number of checks = 256 performs comparable to the CoverBLIP algorithm. However, for improving the overall estimation accuracy if we wish to not use subspace compression, then KDBLIP's time-accuracy performance fails to catch up with that of CoverBLIP. Figure 3 shows the gap between performances of these two algorithms caused by the non-scalability of the KD-tree searches. For instance CoverBLIP with $\varepsilon = 0.4$ outputs more accurate parameter maps (Figure 4(b)) whilst reporting 6x less total search cost.

5.2. In-vivo data with variable-density spiral acquisition

In this part we evaluate reconstruction methods in Table 1 on *in-vivo* MRF data acquired from a healthy volunteer using the IR-BSSFP sequence and the 1.5 T whole body Espree Siemens Healthcare scanner with 32-channel head receiver coil. This dataset was used in the seminal paper of Ma *et al.* [8], where FAs have a pseudo-randomized (Perlin noise) pattern of length $L = 1000$ and TRs are uniformly selected at random between 10.5 and 14 msec. At each time frame (repetition time) one interleaf of the variable-density spiral readout samples the k-space (see [8, Figure 1] for the FA,TR and spiral readout patterns used in this experiment). The spiral trajectory Ω_t rotates by 7.5° in the next time frame to sample different k-space locations and so on. The overall k-space undersampling factor is 48x folds and since a non-Cartesian readout pattern has been used, the operator F in the forward model (1) is implemented using



(a) T2 Map reconstructed by TM i.e. the first iteration of BLIP

(b) T2 Map reconstructed by BLIP

(c) Measurements fidelity error decay through iterations

Figure 5: Reconstruction at $n = 128 \times 128$ spatial resolution (32-coil data). Despite a better data consistency, the iterative scheme (BLIP) reports high-frequency artefacts in the reconstructed maps indicating the lack of sufficient high-resolution information in an spiral readout.

the non-uniform Fourier transform (NUFFT) [38]. Sensitivity maps (i.e. S operator in (1)) are computed off-line from the acquired multi-coil data [36]. For reconstruction a dictionary of $d = 363624$ fingerprints is simulated for combinations of discrete parameters $T1 = [100 : 20 : 2000, 2300 : 300 : 5000]$ (msec), $T2 = [20 : 5 : 100, 110 : 10 : 200, 300 : 200 : 1900]$ (msec), and $B0 = [-250 : 20 : -190, -50 : 1 : 50, 190 : 20 : 250]$ (Hz). As a common practice used to precondition non-Cartesian MRF problems, we incorporate a density compensation scheme within the reconstruction pipeline to enable faster convergence (see more details in the supplementary materials). With this update, we initialize the step-size by the compression factor $\mu = n/m$ similar to the Cartesian sampling. We empirically observe that this choice satisfies the criteria (17) for most of the iterations and for the rest one or two shrinkage sub-iterations suffices.

5.2.1. Missing high-resolution information and high-frequency artefacts As can be observed in Figure 5(b), using iterative methods for spiral readouts may cause high-frequency artefacts in the estimated maps. We would like to emphasize that this issue does not arise because of the deficiency of iterations. Indeed, the monotone decay of the measurement fidelity error implies that iterations improve data consistency as compared to the non-iterative TM scheme (Figure 5(c)). As also highlighted in [28], after an initial rapid decay in the fidelity error a long epoch of slowly-decaying iterations will follow to recover high-resolution image features. However, since spiral readouts do not (sufficiently) sample high-frequency k-space locations, solving (5) may admit undesirable solutions with high-frequency artefacts which appear in the second epoch of iterations until convergence. These artefacts can be removed by either using a spatial-smoothing regularization[¶] or by reconstructing images in a lower spatial resolution. Here we take

[¶] In [28] a low-pass filtering is used at each iteration to remove high-frequency artefacts. However, in conjunction with a non-convex matching step, such a sequential projection approach (i.e. for multiple

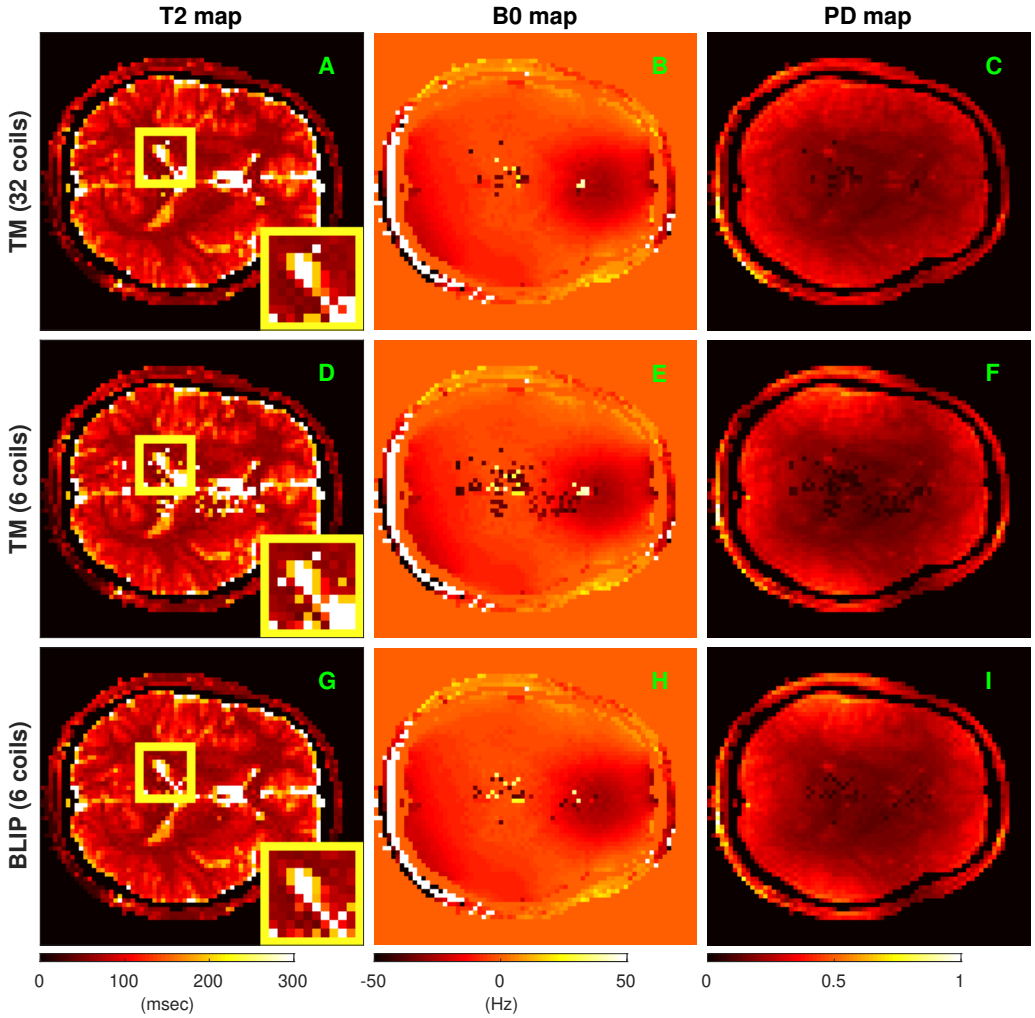


Figure 6: Recovered $T1, T2, B0$ and PD maps from the 32-coil data using TM algorithm (A-D), and those reconstructed from only 6-coil data by using TM (E-H) and iterative BLIP (I-L) algorithms.

the latter approach and reconstruct volunteer images in $n = 100 \times 100$ resolution for the rest of our experiments —instead of the 128×128 resolution maps shown in [8, Figure 3] using the non-iterative TM. We also observe that with this update we require less iterations to converge.

5.2.2. Results The non-iterative TM algorithm performs reasonably well when all 32 coil/channel data are used (Figure 6(A-C)), supporting the fact that in data-rich regimes we may not need sophisticated inference algorithms [57]. To better highlight the advantage of iterations we select measurements from 6 coils that maximally cover the k-space. The recovered PD maps from the 6-coil data (Figure 6(F,I)) demonstrate weaker signal intensity in central and certain border regions as compared to the one obtained from the 32-coil data in Figure 6(C). Comparing Figure 6(A,B) to Figure 6(D,E) shows that TM reconstruction for the reduced 6-coil data introduces artefacts in both $T2$ and

constraints) would not guarantee the convergence of iterations.

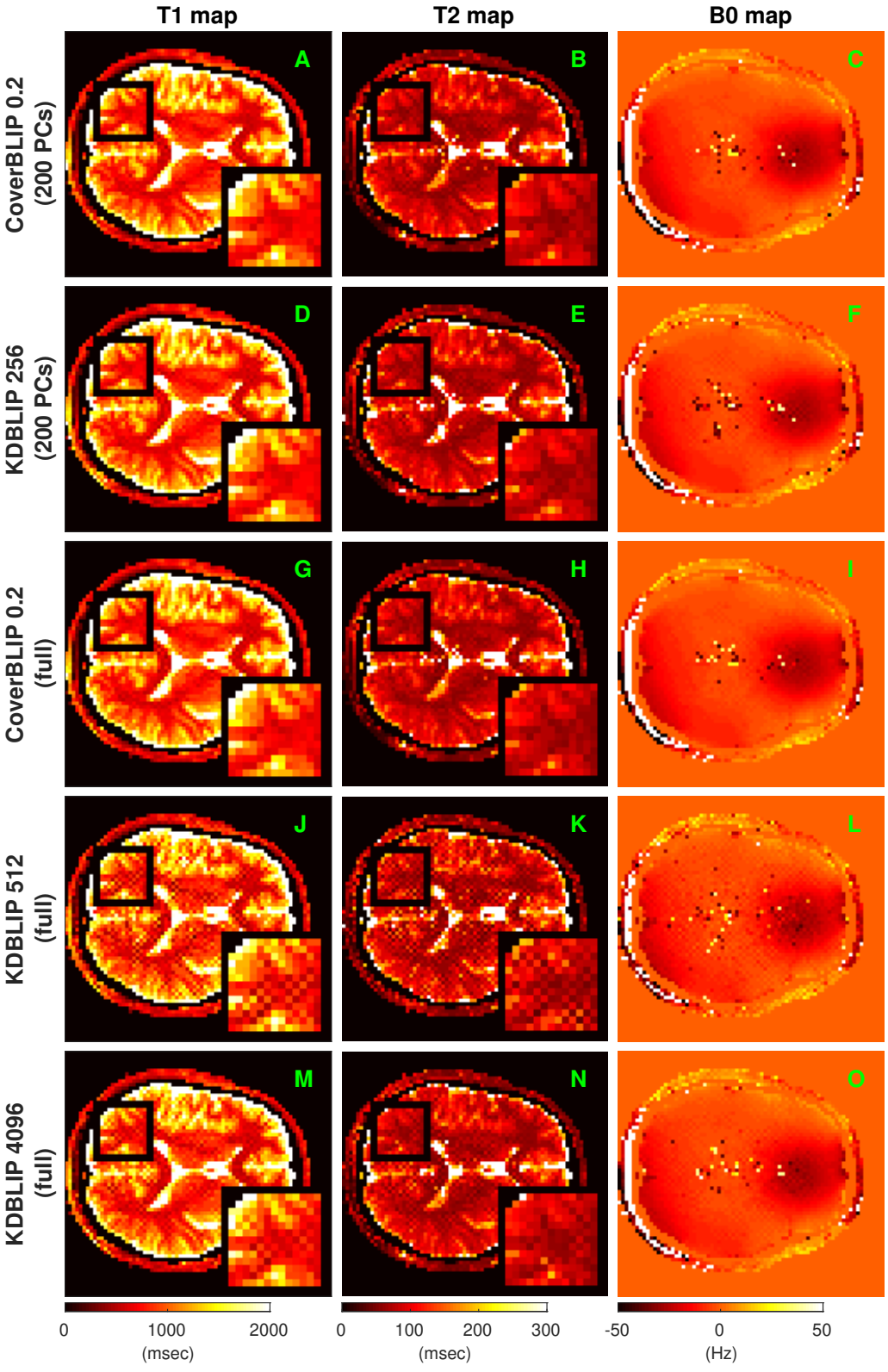


Figure 7: Reconstructed maps using inexact iterations of CoverBLIP and KDBLIP algorithms (6-coil data), tested for different search-dimension regimes i.e. with/without using temporal subspace compression.

Algorithm	BLIP	CoverBLIP			KDBLIP		
ANNS parameter	-	0.1	0.2	0.4	256	512	4096
				Temporal compression $s = 200$			
Total search cost	6.55e12	6.97e10	2.28e10	9.91e9	4.88e10	9.68e10	6.76e11
Reconstruction time (sec)	411.5	192.3	147.2	140.6	153.1	156.5	171.8
Normalized fidelity error	1.285e-1	1.305e-1	1.306e-1	1.309e-1	1.337e-1	1.333e-1	1.328e-1
				No temporal compression			
Total search cost	2.91e13	3.84e11	1.37e11	6.76e10	5.60e11	1.33e12	8.11e12
Reconstruction time (sec)	1660.1	1014.7	684.1	652.4	680.0	825.6	846.6
Normalized fidelity error	1.287e-1	1.307e-1	1.311e-1	1.320e-1	1.490e-1	1.458e-1	1.414e-1

Table 3: Comparison between iterative methods with/without using temporal subspace compression in terms of search cost, total runtime and normalized fidelity error, tested on in-vivo data.

$B0$ maps around the Cerebrospinal Fluid (CSF) regions where the signal is weak.[+] The iterative BLIP algorithm however corrects for this issue and works stably in low-data regime. The rest of our experiments focuses on the 6-coil k-space data.

Table 3 compares the reconstruction performance of iterative methods BLIP, KDBLIP and CoverBLIP for different search dimensions with/without using subspace compression. For subspace compression we use $s = 200$ principal components as recommended in [11]. The corresponding reconstructed maps can be also visually compared in Figure 7. The BLIP algorithm using exact brute-force searches achieves the lowest fidelity error but it requires the longest reconstruction time and highest search complexity. CoverBLIP with $\varepsilon = 0.2$ reports the best reconstruction time-accuracy (also total search cost) among all tested methods. CoverBLIP saves more than 2 orders of magnitude in total search cost of BLIP with a comparable reconstruction accuracy (see the corresponding normalized fidelity errors in Table 3 and the recovered maps in Figure 7(A-C, G-I)). Importantly and unlike KDBLIP, this computational advantage is consistent for all tested search dimensions. We can observe in Figure 7(D-F) that by using subspace compression $s = 200$, the KDBLIP algorithm with 256 checks outputs comparable parameter maps to that of CoverBLIP, however with 2-4 times more search cost. Runtimes reported for both methods in this case (Table 3) are however similar because as previously pointed out we do not claim an optimal implementation of the cover tree searches used here. KDBLIP uses non-scalable tree searches and therefore without a dimensionality reduction —even with a large number of checks = 4096, a longer runtime and 80x higher search cost than CoverBLIP—this algorithm fails to output artefact-free parameter maps (Figure 7(J-O)). More artefacts occur using smaller checks e.g. 512 or 256. In this experiment a moderate subspace compression turns out to be advantageous for all tested algorithms, but then it is a crucial step for using KD-tree searches. We empirically observed that KDBLIP starts reporting poor reconstruction

[+] $T1$ maps were identical and therefore now shown here.

time-accuracies when more than 350 principal components are used.

Comparing the overall runtimes in Table 3, we note that CoverBLIP ($\varepsilon = 0.2$) achieves 2.5-3x fold acceleration compared to the BLIP algorithm which is less than what was reported for our previous synthetic data experiment in Section 5.1. The reason is that here we use multi-coil data and non-Cartesian k-space sampling where both make the gradient updates become a non-trivial computational overhead for the iterations. Note that reconstructions from a non-Cartesian acquisition protocol requires computing slow NUFFT operations in each iteration. As a result, despite a significant reduction in the total search cost (i.e. projection steps) this advantage will be less pronounced in the overall runtime of CoverBLIP. We believe addressing this issue i.e. breaking down the cost of heavy gradient updates, merits an independent line of future investigation beyond the scope of this work.

6. Conclusions and future directions

We considered accelerating the iterative scheme for model-based MRF reconstruction and for this purpose we approximated the matched-filtering step in each iteration using cover tree's $(1+\varepsilon)$-ANNS search scheme. For low-dimensional manifold datasets cover trees offer appealing construction times, memory requirements and remarkably low search complexities scaling logarithmic in terms of data population. With this motivation, we proposed the CoverBLIP algorithm which adopts such tree structures for fast iterative searches over large-size MRF dictionaries i.e. discrete manifold of Bloch responses parametrized by few NMR characteristics. Provided with a notion of (model-restricted) embedding we showed that the inexact iterations of CoverBLIP linearly convergence toward a solution with the same order of accuracy as when using BLIP with exact brute-force searches. We also introduced an adaptive step-size scheme that guarantees local monotone convergence of CoverBLIP in the absence of bi-Lipschitz embedding. We evaluated the performance of our proposed method on both synthetic and real-world MRF datasets using different sampling strategies, and we demonstrated that CoverBLIP is capable of achieving orders of magnitude acceleration in conducting the projection steps as compared to the exact iterations of BLIP. We also showed that CoverBLIP is a scalable algorithm able to maintain the gain in its time-accuracy performance in high-dimensional search spaces.

Future works include application of CoverBLIP to the emerging multi-parametric MRF problems with more complex dynamic responses encoding a larger number of NMR characteristics. In such cases and due to the inherent non-linearity of Bloch responses a low-dimensional subspace model of the dictionary would be prohibitively inaccurate, and one would rather need to resort on fast search schemes such as cover trees that are robust against the curse-of-dimensionality. Our current search implementation does not benefit from the considerable amount of inter-voxel correlations present in a *query batch*. As shown in e.g. [58] faster searches are possible by additionally building a dual (cover) tree on the query batches. An interesting line of future work would adopt

this idea to further accelerate CoverBLIP, however with more restricted choice of dual trees or batch sizes whose construction times would not bring a computational overhead throughout multiple iterations. Further extension to the present work could also focus on reducing the computational cost of the gradient updates for non-Cartesian and multi-coil acquisition schemes. In this regard, a possible line of investigation would be the application of *randomized* iterative projected gradient algorithms (see e.g. [59, 60]), where iterations adopt cheap, unbiased and variance-reduced stochastic approximations of the true but computationally-intensive gradient updates.

Acknowledgments

We thank Dr. Dan Ma (Case Western Reserve University) for providing us the real data used in [8]. We also thank Dr. Pedro Gòmez and Dr. Marion Menzel for useful discussions during MG's visit to GE Healthcare in Munich. This work is partly funded by the EPSRC grant EP/M019802/1 and the ERC C-SENSE project (ERCADG- 2015-694888). MG is also supported by the Scottish Research Partnership in engineering (SPRe) Award PECRE1718/18.

References

[1] Golbabaee M, Chen Z, Wiaux Y and Davies M E 2017 Cover tree compressed sensing for fast MR fingerprint recovery *2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP)* pp 1–6

[2] Golbabaee M, Chen Z, Wiaux Y and Davies M E 2018 CoverBLIP: scalable iterative matched-filtering for MR fingerprint recovery *Proceedings of the Joint Annual Meeting ISMRM-ESMRMB, Paris, France*

[3] Tofts P 2005 *Quantitative MRI of the brain: measuring changes caused by disease* (John Wiley & Sons)

[4] Look D C and Locker D R 1970 *Review of Scientific Instruments* **41** 250–251

[5] Homer J and Beevers M S 1985 *Journal of Magnetic Resonance (1969)* **63** 287–297

[6] Huang T Y, Liu Y J, Stemmer A and Poncelet B P 2007 *Magnetic resonance in medicine* **57** 960–966

[7] Deoni S C, Peters T M and Rutt B K 2005 *Magnetic resonance in medicine* **53** 237–241

[8] Ma D, Gulani V, Seiberlich N, Liu K, Sunshine J, Durek J and Griswold M 2013 *Nature* **495** 187–192

[9] Jaynes E 1955 *Physical Review* **98** 1099

[10] Jiang Y N, Ma D, Seiberlich N, Gulani V and Griswold M 2015 *Magnetic resonance in medicine* **74** 1621–1631

[11] McGivney D F, Pierre E, Ma D, Jiang Y, Saybasili H, Gulani V and Griswold M A 2014 *IEEE transactions on medical imaging* **33** 2311–2322

[12] Davies M, Puy G, Vandergheynst P and Wiaux Y 2014 *SIAM Journal on Imaging Sciences* **7** 2623–2656

[13] Zhao B, Setsompop K, Adalsteinsson E, Gagoski B, Ye H, Ma D, Jiang Y, Ellen Grant P, Griswold M A and Wald L L 2018 *Magnetic resonance in medicine* **79** 933–942

[14] Assländer J, Cloos M A, Knoll F, Sodickson D K, Hennig J and Lattanzi R 2018 *Magnetic resonance in medicine* **79** 83–96

[15] Wang C, Coppo S, Mehta B, Seiberlich N, Yu X and Griswold M 2017 Magnetic resonance fingerprinting with quadratic rf phase for simultaneous measurement of δf, T1, T2, and T2* *Proceedings of the Annual Meeting of ISMRM, Honolulu*

[16] Rieger B, Zimmer F, Zapp J, Weingärtner S and Schad L R 2017 *Magnetic resonance in medicine* **78** 1724–1733

[17] Su P, Mao D, Liu P, Li Y, Pinho M C, Welch B G and Lu H 2017 *Magnetic resonance in medicine* **78** 1812–1823

[18] Wright K L, Jiang Y, Ma D, Noll D C, Griswold M A, Gulani V and Hernandez-Garcia L 2018 *Magnetic resonance imaging* **50** 68–77

[19] Yu A C, Badve C, Ponsky L E, Pahwa S, Dastmalchian S, Rogers M, Jiang Y, Margevicius S, Schluchter M, Tabayoyong W *et al.* 2017 *Radiology* **283** 729–738

[20] Lemasson B, Pannetier N, Coquery N, Boisserand L S, Collomb N, Schuff N, Moseley M, Zaharchuk G, Barbier E and Christen T 2016 *Scientific reports* **6** 37071

[21] Beygelzimer A, Kakade S and Langford J 2006 Cover trees for nearest neighbor *Proceedings of the 23rd international conference on Machine learning* (ACM) pp 97–104

[22] Donoho D L 2006 *IEEE Transactions on Information Theory* **52** 1289–1306 ISSN 0018-9448

[23] Baraniuk R G and Wakin M B 2009 *Foundations of Computational Mathematics* **9** 51–77 ISSN 1615-3383

[24] Baraniuk R G, Cevher V, Duarte M F and Hegde C 2010 *IEEE Transactions on Information Theory* **56** 1982–2001 ISSN 0018-9448

[25] Blumensath T 2011 *IEEE Transactions on Information Theory* **57** 4660–4671 ISSN 0018-9448

[26] Golbabaee M and Davies M E 2018 *IEEE Transactions on Information Theory* **64** 6707–6721

[27] Cauley S F, Setsompop K, Ma D, Jiang Y, Ye H, Adalsteinsson E, Griswold M A and Wald L L 2015 *Magnetic resonance in medicine* **74** 523–528

[28] Cline C C, Chen X, Mailhe B, Wang Q, Pfeuffer J, Nittka M, Griswold M A, Speier P and Nadar M S 2017 *Magnetic resonance imaging* **41** 29–40

[29] Krauthgamer R and Lee J R 2004 Navigating nets: Simple algorithms for proximity search *Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms* SODA '04

[30] Zhao B 2015 Model-based iterative reconstruction for magnetic resonance fingerprinting *Image Processing (ICIP), IEEE International Conference on* pp 3392–3396

[31] Mazor G, Weizman L, Tal A and Eldar Y C 2016 Low rank magnetic resonance fingerprinting *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the* (IEEE) pp 439–442

[32] Doneva M, Amthor T, Koken P, Sommer K and Börnert P 2017 *Magnetic resonance imaging* **41** 41–52

[33] Cohen O, Zhu B and Rosen M S 2018 *Magnetic resonance in medicine* **80** 885–894

[34] Virtue P, Yu S X and Lustig M 2017 *arXiv preprint arXiv:1707.00070*

[35] Golbabaee M, Chen D, Gómez P A, Menzel M I and Davies M E 2019 Geometry of deep learning for magnetic resonance fingerprinting *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* pp 7825–7829 ISSN 2379-190X

[36] Walsh D O, Gmitro A F and Marcellin M W 2000 *Magnetic Resonance in Medicine* **43** 682–690

[37] Benjamin A, Gómez P, Golbabaee M, Sperenger T, Menzel M, Davies M and Marshall I 2018 Balanced multi-shot EPI for accelerated cartesian MR fingerprinting: An alternative to spiral MR fingerprinting *Proceedings of the Joint Annual Meeting ISMRM-ESMRMB, Paris, France*

[38] Fessler J A and Sutton B P 2003 *IEEE Transactions on Signal Processing* **51** 560–574

[39] Tang S, Fernandez-Granda C, Lannuzel S, Bernstein B, Lattanzi R, Cloos M, Knoll F and Assländer J 2018 *Inverse problems* **34** 094005

[40] Elhamifar E and Vidal R 2013 *IEEE transactions on pattern analysis and machine intelligence* **35** 2765–2781

[41] Golbabaee M and Vandergheynst P 2012 Hyperspectral image compressed sensing via low-rank and joint-sparse matrix recovery *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on* (IEEE) pp 2741–2744

[42] Golbabaee M, Arberet S and Vandergheynst P 2013 *IEEE Transactions on Image Processing* **22** 5096–5110 ISSN 1057-7149

[43] Daubechies I, Defrise M and De Mol C 2004 *Communications on pure and applied mathematics* **57** 1413–1457

[44] Beck A and Teboulle M 2009 *SIAM journal on imaging sciences* **2** 183–202

[45] Blumensath T and Davies M E 2009 *Applied and Computational Harmonic Analysis* **27** 265 – 274 ISSN 1063-5203

[46] Shah P and Chandrasekaran V 2011 Iterative projections for signal identification on manifolds: Global recovery guarantees *2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)* pp 760–767

[47] Bentley J L 1975 *Communications of the ACM* **18** 509–517

[48] Guttman A 1984 *R-trees: A dynamic index structure for spatial searching* vol 14 (ACM)

[49] Omohundro S M 1989 *Five balltree construction algorithms* (International Computer Science Institute Berkeley)

[50] Indyk P and Motwani R 1998 Approximate nearest neighbors: towards removing the curse of dimensionality *Proceedings of the thirtieth annual ACM symposium on Theory of computing* (ACM) pp 604–613

[51] Izbicki M and Shelton C 2015 Faster cover trees *International Conference on Machine Learning* pp 1162–1170

[52] Blumensath T and Davies M E 2010 *IEEE Journal of selected topics in signal processing* **4** 298–309

[53] Baraniuk R, Davenport M, DeVore R and Wakin M 2008 *Constructive Approximation* **28** 253–263

[54] `http://www.cs.ubc.ca/research/flann/`

[55] `http://hunch.net/~jl/projects/cover_tree/cover_tree.html`

[56] `http://brainweb.bic.mni.mcgill.ca/brainweb/`

[57] Chandrasekaran V and Jordan M I 2013 *Proceedings of the National Academy of Sciences* **110** E1181–E1190

[58] Curtin R R 2015 *Improving dual-tree algorithms* Ph.D. thesis Georgia Institute of Technology

[59] Bottou L 2010 Large-scale machine learning with stochastic gradient descent *Proceedings of COMPSTAT* (Springer) pp 177–186

[60] Tang J, Golbabaee M and Davies M 2017 *Proceedings of the 34th International Conference on Machine Learning* **70** 3377–3386