Designing and learning non-Cartesian k-space sampling trajectories for accelerated 3D MRI

Chaithya Giliyar Radhakrishna

université
PARIS-SACLAY

Designing and learning non-Cartesian k-space sampling trajectories for accelerated 3D MRI

Conception et apprentissage de trajectoires d'échantillonnage non cartésiennes dans l'espace k pour l'IRM 3D accélérée

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 575, Electrical, Optical, Bio : physics and Engineering (EOBE)
Spécialité de doctorat : Physique et imagerie médicale
Graduate School : Sciences de l'ingénierie et des systèmes
Référent : Faculté des sciences d'Orsay

Thèse préparée dans l'unité de recherche **NeuroSpin** (Université Paris-Saclay, CEA), sous la direction de **Philippe CIUCIU**, Directeur de recherche

Thèse soutenue à Paris-Saclay, le 25 avril 2023, par

Chaithya GILIYAR RADHAKRISHNA

Composition du jury

Membres du jury avec voix délibérative

Charles SOUSSEN Professeur, CentraleSupélec	Président
Daniel RUECKERT Professeur, Technical University of Munich	Rapporteur & Examinateur
Douglas NOLL Professeur, University of Michigan	Rapporteur & Examinateur
Myriam EDJLALI Professeur, Assistance Publique-Hôpitaux de Paris	Examinatrice
Ludovic de ROCHEFORT Chargé de recherche, Université Aix-Marseille	Examinateur

THESE DE DOCTORAT
NNT : 2023UPAST049

université PARIS-SACLAY

ÉCOLE DOCTORALE
Physique et ingénierie : Electrons, Photons, Sciences du vivant (EOBE)

Title : Designing and learning non-Cartesian k-space sampling trajectories for accelerated 3D MRI
Keywords : MR image reconstruction ; susceptibility-weighted imaging ; non-Cartesian imaging ; off-resonance correction ; deep learning ; SPARKLING trajectories.

Abstract : Magnetic resonance Imaging (MRI) is a popular non-invasive medical imaging technique to understand structures of tissues in the human body. However, its reach and use is strongly limited by long acquisition times, as the data acquisition is segmented in time and performed in the Fourier domain of the image or k-space. Compressed sensing theories enabled shorter scan times by sensing less compressible information, particularly by adopting a variable density sampling (VDS) in k-space : i.e. the lower frequencies (center of k-space) have to be sampled more densely than the higher at its periphery. Efficient ways to achieve such VDS is through the use of non-Cartesian sampling, where k-space is sampled along curves rather than conventional gridded lines. Essentially, this involves the optimization of 1) the sampling pattern in k-space under MR hardware constraints and 2) image reconstruction from undersampled k-space data.

In this thesis, we address the above issues in the context of accelerated 3D MRI. Firstly, we extend the earlier developed SPARKLING algorithm to 3D, which results in hardware compliant non-Cartesian trajectories that meet a user specified target sampling density (TSD) in k-space. Secondly, we tackle the issue of off-resonance artifacts in these SPARKLING trajectories through a modification of the underlying algorithm, to allow for signal recovery at regions where the off-resonance is high. Thirdly, we shift to data-driven learning based approach to learn the TSD for SPARKLING trajectories. Last but not least, we propose a novel deep learning based approach to learn hardware compliant k-space sampling trajectories jointly with a corresponding image reconstruction network. Overall, our efforts in this thesis allowed for an acceleration of up to 20x in 3D MRI scans with a significant reduction in off-resonance artifacts, allowing to image with minimum degradation in reconstructed image quality.

Titre : Conception et apprentissage de trajectoires d'échantillonnage non cartésiennes dans l'espace k pour l'IRM 3D accélérée
Mots clés : reconstruction d'image IRM ; imagerie pondérée en susceptibilité magnétique ; acquisitions non-cartésiennes ; correction d'effet hors-resonance ; apprentissage profond ; trajectoires SPARKLING.

Résumé : L'imagerie par résonance magnétique (IRM) est une technique d'imagerie médicale non invasive très répandue qui permet de comprendre les structures des tissus du corps humain. Cependant, sa portée et son utilisation sont fortement limitées par les longs temps d'acquisition, car l'acquisition des données est segmentée en temps et réalisée dans le domaine de Fourier de l'image ou espace k. Les théories d'échantillonnage compressif ont permis d'accéder à des temps d'examen plus courts en collectant moins d'information compressible, en particulier en adoptant dun échantillonnage à densité variable (EDV) dans l'espace k : les basses fréquences (au centre de l'espace k) doivent être échantillonnées plus densément que les hautes fréquences situées à la périphérie de celui-ci. Un moyen efficace d'obtenir un tel EDV est d'utiliser un échantillonnage non cartésien, où l'espace k est échantillonneur le long de courbes plutôt que de lignes alignées sur une grille cartésienne. Essentiellement, cela implique l'optimisation 1) du modèle d'échantillonnage dans l'espace k sous les contraintes matérielles de l'IRM et 2) la reconstruction de l'image à partir des données sous-échantillonnées de l'espace k.

Dans cette thèse, nous abordons les questions ci-dessus dans le contexte de l'IRM 3D accélérée. Tout d'abord, nous étendons l'algorithme SPARKLING développé précédemment à l'imagerie 3D, ce qui permet d'obtenir des trajectoires non cartésiennes conformes aux contraintes matérielles et fidèles à une densité d'échantillonnage cible (DEC) spécifiée par l'utilisateur dans l'espace k. Ensuite, nous abordons le problème des artefacts hors résonance présents au sein des trajectoires SPARKLING en modifiant l'algorithme sous-jacent, afin de permettre la récupération du signal dans les régions où la hors résonance est élevée. Dans la dernière partie de la thèse, nous passons à une approche basée sur l'apprentissage piloté par les données pour apprendre la DEC pour les trajectoires SPARKLING. Enfin, nous proposons une nouvelle approche basée sur l'apprentissage profond pour apprendre des courbes d'échantillonnage de l'espace k conformes aux contraintes matérielles conjointement avec un réseau de reconstruction d'images adapté. Dans l'ensemble, nos efforts dans cette thèse ont permis d'accélérer jusqu'à 20 fois les scans IRM 3D avec une réduction significative des artefacts hors résonance, permettant ainsi une dégradation minimale de la qualité des images reconstruites.

विद्या ददाति विनयं विनयाद्याति पात्रताम् ।
पात्रत्वाद्धनमाप्नोति धनाद्धर्मं ततः सुखम् ॥ ५ ॥

knowledge gives humility & discipline,
from these comes worthiness to render good deeds,
& from that comes joy!

— Saral Hitopadesha, Sanskrit Fables

Acknowledgements

I would like to express my heartfelt gratitude to everyone who have played pivotal roles in the completion of this thesis. First and foremost, I extend my deepest appreciation to my parents for their unwavering support, encouragement, and belief in my abilities throughout this journey. I am profoundly indebted to my supervisor, Philippe Ciuciu, for his invaluable guidance, mentorship, and dedication to my academic growth. I extend my sincere thanks to Alexandre Vignaud and Frank Mauconduit for their exceptional mentoring and continuous assistance during the course of my research. I would like to acknowledge the collaborative efforts of my fellow PhD students, Guillaume Daval-Frérot and Zaccharie Ramzi, whose insights and contributions greatly enriched this work. Special thanks go to my fellow PhD students, including Zaineb Amor, Pierre-Antoine Comby, and Redouane Jamil, for their camaraderie and support. I am immensely grateful to the entire MIND and METRIC teams, as well as the Neurospin Institute in general, for providing a stimulating and nurturing research environment. Lastly, but certainly not the least, I want to express my deep appreciation to my wife for her unwavering support, understanding, and love throughout this journey. Each of you has played an integral role in shaping this thesis, and I am sincerely thankful for your contributions.

Funding Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 800945-NUMERICS-H2020-MSCA-COFUND-2017. This work was granted access to the CCRT High-Performance Computing (HPC) facility at CEA under the Grant CCRT2022-gilirach and the HPC resources of IDRIS under the allocation 2021-AD011011153 made by GENCI.

* * *
* *
*

Contents

General Introduction

Context & Motivations

The quest for efficient sampling strategies has been a major challenge in Magnetic Resonance Imaging (MRI) since its invention. The theory of Compressed Sensing (CS) [Lus+07] boosted this quest by providing significant theoretical insights. It was proved and observed empirically that for under-sampled acquisitions and compressible signals in an orthogonal basis, an efficient implementation relies on trajectories with a variable density in k-space: The lower frequencies located in the center of k-space have to be sampled more densely than the higher in its periphery [Puy+11; Cha+14; Adc+17; Boy+19]. Non-Cartesian k-space trajectories (e.g. spiral, radial, rosette, etc.) [Ahn+86a; Mey+92; Jac+92; Nol97; Law+09; Lus+05a] have been proposed for accelerated and robust-to-motion 2D imaging, prior to the existence of theoretical CS foundations. While being compliant with scanner hardware constraints on the gradients, these trajectories do not sample the k-space according to a well controlled TSD. For instance, in spiral imaging, fulfilling these constraints transforms an initially prescribed density into another one [Cha15, p. 97]. Recently, the SPARKLING algorithm [Boy+16; Cha+17; Laz+19] has been shown to automatically generate optimized non-Cartesian sampling patterns compatible with MR hardware constraints on maximum gradient amplitude and slew rate. SPARKLING optimally samples the k-space (see [Puy+11; Cha+14]) with a controlled distribution of samples (e.g., Variable Density Sampling (VDS)) and a locally uniform k-space coverage.

Contributions

The main focus and goal of this thesis involves designing and learning 3D k-space sampling trajectories for minimally artifacted and accelerated acquisitions. In this regard, we initially focus on extending and improving the SPARKLING trajectories, and later use foundations of deep learning to learn hardware compliant k-space sampling trajectories. The main contributions and goals of this PhD thesis can be sorted out into the following categories:

Towards fully 3D SPARKLING algorithm [Cha+22d]

Due to Signal-to-Noise Ratio (SNR), 3D imaging is preferred to reach isotropic high-resolution imaging (e.g. 600µm isotropic). In this regard, multiple approaches have been utilized to efficiently down-sample 3D k-space. Some of these involve a combination of a readout in the z-direction with a 2D under-sampled mask based on Poisson disk sampling [Vas+10]. Additional attempts on full 3D readouts were proposed like 3D radial trajectory [Lar+08], 3D cones [Ira+95], twisted projections [Boa+97] and hybrid radial-cones [Joh17]. However, these trajectories were primarily based on parameterizing a k-space sampling curve, and the final sampling pattern was not optimized with respect to image reconstruction quality. Some recent studies explored how to optimize the sampling pattern [Dal+04; Mir+04; Kum+08], but did not include a clear sampling criterion in order to maximize the image reconstruction quality. While methods like [HK19; See+10] do use Cramér-Rao bound for sparse signals [BE10] as a tailored optimality criterion, they are computationally demanding thereby limiting its use to the Cartesian setting. Other methodologies in the literature involved stacking a 2D under-sampled trajectory like stack of stars [Son+04; Lo08], stack of

spirals [Ira+95; The+99] and stack of 2D SPARKLING [Laz+20a]. Further in [Laz+20a], a local 3D SPARKLING approach was proposed by designing a single trajectory within a cone obtained from a parcellation of the 3D spherical k-space. Then all the cones covering a given elevation plane were filled up using the replication of the same trajectory. However, this method did not ensure a locally uniform sampling pattern at the boundaries of cones as the problem was only solved locally.

Solving the SPARKLING problem globally in 3D is computationally expensive with $O(N^2)$ order of computational complexity, where N represents the number of sampling locations in the pattern. We proposed an implementation based on fast multipole methods (FMM) [Wan+19; Fon+09a] that allowed us to design sampling patterns with up to 10^7 k-space samples, thus opening the door to 3D VDS. We compared multi-CPU and GPU implementations and demonstrated that the latter is optimal for 3D imaging in the high-resolution acquisition regime (600µm isotropic). Finally, we showed that this novel optimization for full 3D SPARKLING outperforms stacking strategies or 3D twisted projection imaging through both retrospective and prospective studies on both a NIST phantom and in vivo brain scans at 3 Tesla. Overall the proposed method allows for 2.5-3.75x shorter scan times compared to GRAPPA-4 parallel imaging acquisition at 3 Tesla without compromising image quality.

Addressing issues in prospective scans for non-Cartesian imaging [Cha+22a; Cha+22c; Cha+22b; Cha+]

After extending the SPARKLING trajectories to 3D, a large gap was observed in the reconstructed image quality between retrospective simulations and actual prospective scans (see [Cha+22d, Fig. 10]) in the case of T_2^*-weighted imaging. Such a discrepancy in prospective setting was identified [Cha+22d, S2.B],[Dav+22] to be due to the presence of a strong off-resonance effects and T_2^* decay whose effects are accumulated in T_2^*-w contrast due to longer Echo Time (TE) used to enhance the susceptibility contribution. This effect is amplified in non-Cartesian imaging and notably in 3D SPARKLING as such trajectories have arbitrary readout directions, leading to local k-space inconsistencies. Although these artifacts can be corrected [Sut+03] without needing any supplementary scan for ΔB_0 map [Dav+22], such corrections are computationally expensive.

Additionally, another limitation of the SPARKLING trajectories is that we use affine TE constraints, where we limit the trajectory to pass through the center of k-space at echo time to obtain images at required target imaging contrast. This results in a strong oversampling of the center of k-space compared to the Nyquist criteria which can be sub-optimal as the extra samples can be assigned to higher frequencies instead, thus resulting in improved image reconstructions with finer details in structures.

In this work, we address both above described issues through modification of the SPARKLING cost function and constraint set. Through MORE-SPARKLING , we introduced a temporal weighting in the traditional SPARKLING optimization problem which takes the temporal nature of the sampled data into account. In GoLF-SPARKLING , we incorporated the features of Cartesian sampling into the SPARKLING framework through more general affine constraints and adaptation of the TSD to match these Nyquist criteria constraints. Prospective k-space data was acquired at 3 Tesla on 20-fold accelerated trajectories as compared to fully sampled Cartesian reference. Through in vivo experiments, MORE-SPARKLING allowed us to recover signal dropouts observed on original SPARKLING acquisitions at larger B_0 field inhomogeneities. Further, through GoLF-SPARKLING i.e. Cartesian sampling at center of k-space provided improved reconstructed image quality with reduced artifacts.

Jointly learning k-space sampling pattern and reconstruction network [CRC21; CRC22; CC22; CC]

A limitation of SPARKLING algorithm is the need to setup a TSD as an input to the algorithm. For this, we learned the sampling density using LOUPE [Bah+20a], and used it as TSD to the SPARKLING algorithm to generate 2D SPARKLING trajectories [CRC21]. However, the gridded TSD was learned in the Cartesian domain, while the actual trajectory

being optimized was non-Cartesian. As this could lead to suboptimal results, there is a need for directly learning hardware compliant k-space sampling trajectories along with an image reconstruction network in a non-Cartesian setting.

In this regard, new methods [Wei+20; Wan+21; Ved+20] have been developed to overcome the need for estimating a TSD, through direct joint learning of the non-Cartesian k-space sampling trajectories and MR image reconstruction in a data-driven manner on the fastMRI dataset [Zbo+18]. In [Wei+20; Ved+20], the authors jointly learned PILOT trajectories along with U-net parameters as a reconstruction model to denoise the basic image yielded by the adjoint of the Nonuniform Fast Fourier Transform (NUFFT) operator. However, this method relies on auto-differentiation of the NUFFT operator, which is inaccurate numerically as observed in [WF23], resulting in sub-optimal local minima. This suboptimality was actually reflected in the final shape of the learned trajectories, which only slightly deviated from their initialization. In BJORK [Wan+21], the authors used [WF23] to obtain a more accurate Jacobian approximation of the NUFFT operator for backpropagation. Both above referenced approaches [Wei+20; Wan+21] enforced the hardware constraints by adding penalty terms to the the loss that is minimized during training. Although a viable option, this requires tuning a hyper-parameter associated with each of these penalty terms in the cost function and it does not guarantee that the optimized trajectories will strictly meet these constraints. Further, these penalty terms affect the overall gradients of the loss function, thereby resulting in suboptimality of the trajectories.

In this work [CC], we first developed a generic model called PROjection for Jointly lEarning non-Cartesian Trajectories while Optimizing Reconstructor (PROJeCTOR). More precisely, we introduced a method that learns the k-space trajectories in a data-driven manner while embedding a projected gradient descent algorithm [Cha+16] to fulfill the hardware constraints during the training stage. Unlike BJORK, we directly learned the k-space sampling trajectories and used multi-resolution [Leb+19] similar to SPARKLING to limit the number of trainable parameters at each step. Then, we compared these PROJeCTOR results to PILOT [Wei+20] and BJORK [Wan+21] in 2D MRI. In 2D retrospective studies, our novel PROJeCTOR trajectories presents an improved image reconstruction quality at a 20-fold acceleration factor on the fastMRI data set with SSIM scores of nearly 0.92-0.95 in as compared to corresponding Cartesian reference and also see 3-4dB gain in PSNR as compared to earlier state-of-the-art methods. In a more controlled setting we showed the importance of the projection step during the optimization of k-space trajectories and demonstrated its superiority over penalty-based methods like PILOT and BJORK to enforce hardware constraints. Finally we extended the algorithm to 3D and by comparing optimization to learning based projection schemes, we brought evidence that data-driven joint learning based PROJeCTOR trajectories outperform model-based methods like SPARKLING through a 2dB gain in PSNR and 0.02 gain in SSIM.

Thesis Outline

Chapter 1: *Introduction to Magnetic Resonance Imaging* introduces to MRI, the source of the measured signal and how it is localized. Later, we discuss on the need for accelerating the Magnetic Resonance (MR) scan and discuss some methods used to accelerate scans in traditional Cartesian MRI.

Chapter 2: *Compressed Sensing and Non-Cartesian MRI* introduces CS based reconstruction of MRI data. Further, we discuss the need to sample k-space with VDS for reaching higher acceleration factors which can be achieved through sampling the k-space along curves using non-Cartesian sampling. For this, we present a generalized framework for sampling the k-space using trajectories and discuss its constraints. We finally present some novel non-Cartesian k-space sampling trajectories proposed in literature.

Chapter 3: *Deep Learning for MRI* focuses on some fundamentals and formalism of DL. We present some popular DL architectures and discuss the methods proposed in literature to reconstruct the MR image from under-sampled k-space data. Further, we also summarize

the literature on learning based techniques to optimize the k-space sampling patterns and trajectories.

We later focus on the main contributions during this thesis. In Chapter 4: *Optimizing full 3D SPARKLING trajectories*, we extend the SPARKLING algorithm to 3D allowing us to accelerate the MRI in all the 3 dimensions. However, strong off-resonance artifacts was observed in prospective studies as these optimized trajectories had arbitrary readout directions.

To tackle this, in Chapter 5: *Reducing artifacts in SPARKLING* we present 2 novel updates to the SPARKLING algorithm: MORE and GoLF.

As the SPARKLING algorithm needs TSD as input, in Chapter 6: *Learning sampling density for 2D SPARKLING* we present 4 different candidate densities and show that data-driven learned LOUPE method provides the best TSD. However, this method involves a gridded TSD, which is learned in Cartesian setting and later used to generate non-Cartesian trajectories. This mismatch in learning domains where training is done in Cartesian setting and testing is done in non-Cartesian setting is a major drawback of this method.

This issue is tackled through direct learning of the k-space trajectories in Chapter 7: *Learning trajectories with reconstructor*, and we emphasize the need for a projection step inspired from SPARKLING to prevent suboptimality.

Finally, in Chapter 8: *Conclusions and Perspectives* we summarize the contributions of this thesis and discuss the future directions and perspectives.

In Appendix A: *3D SPARKLING* and Appendix B: *MORE and GoLF* we present some additional theoretical and experimental results to complete the contributions in Chapter 4 and Chapter 5 respectively.

Finally, in Appendix C: *Software and Open Source Contributions* I discuss some software packages developed during the course of the thesis and some open source contributions.

* * *
* *
*

Résumé étendu en français

Contexte & motivations

Depuis son invention, la recherche de stratégies d'échantillonnage efficaces a constitué un défi majeur pour l'imagerie par résonance magnétique (IRM). La théorie du compressed sensing (CS) [Lus+07] a donné un coup de fouet à cette quête en apportant des éclaircissements théoriques significatifs. Il a été prouvé et observé empiriquement que pour des acquisitions sous-échantillonnées et des signaux compressibles dans une base orthogonale, une mise en œuvre efficace repose sur des trajectoires à densité variable dans l'espace k : Les fréquences les plus basses situées au centre de l'espace-k doivent être échantillonnées plus densément que les plus hautes à sa périphérie [Puy+11; Cha+14; Adc+17; Boy+19].

Des trajectoires non cartésiennes dans l'espace k (par exemple spirales, radiales, rosettes, etc.) [Ahn+86a; Mey+92; Jac+92; Nol97; Law+09; Lus+05a] ont été proposées pour l'imagerie 2D accélérée et robuste au mouvement, avant l'existence des fondements théoriques de la CS. Bien qu'elles soient conformes aux contraintes matérielles du scanner sur les gradients, ces trajectoires n'échantillonnent pas l'espace k selon un Target Sampling Density (TSD) bien contrôlé. Par exemple, dans l'imagerie spirale, le respect de ces contraintes transforme une densité initialement prescrite en une autre [Cha15, p. 97]. Récemment, l'algorithme Spreading Projection Algorithm for Rapid K-space sampLING (SPARKLING) [Boy+16; Cha+17; Laz+19] s'est avéré capable de générer automatiquement des modèles d'échantillonnage non cartésiens optimisés compatibles avec les contraintes du matériel MR sur l'amplitude maximale du gradient et la vitesse de balayage. SPARKLING échantillonne de manière optimale l'espace k (voir [Puy+11; Cha+14]) avec une distribution contrôlée des échantillons (par exemple, Variable Density Sampling (VDS)) et une couverture localement uniforme de l'espace k.

Contributions

L'objectif principal de cette thèse est de concevoir et d'apprendre des trajectoires d'échantillonnage 3D dans l'espace k pour minimiser les artefacts et accélérer les acquisitions. A cet égard, nous nous concentrons d'abord sur l'extension et l'amélioration des trajectoires SPARKLING , puis nous utilisons les fondements de l'apprentissage profond pour apprendre des trajectoires d'échantillonnage de l'espace k compatibles avec le matériel. Les principales contributions et les objectifs de cette thèse de doctorat peuvent être classés dans les catégories suivantes :

Vers un algorithme d'étincelle entièrement 3D [Cha+22d]

En raison de Signal-to-Noise Ratio (SNR), l'imagerie 3D est préférée pour obtenir une imagerie isotrope à haute résolution (par exemple 600µm isotrope). À cet égard, de nombreuses approches ont été utilisées pour réduire efficacement l'échantillonnage de l'espace k en 3D. Certaines d'entre elles impliquent une combinaison d'une lecture dans la direction z avec un masque sous-échantillonné en 2D basé sur l'échantillonnage du disque de Poisson [Vas+10]. D'autres tentatives de lectures 3D complètes ont été proposées, comme la trajectoire radiale 3D [Lar+08], les cônes 3D [Ira+95], les projections torsadées [Boa+97] et les cônes radiaux hybrides [Joh17]. Cependant, ces trajectoires étaient principalement basées sur la paramétrisation d'une courbe d'échantillonnage dans l'espace k, et le motif d'échantillonnage final n'était

pas optimisé en ce qui concerne la qualité de la reconstruction de l'image. Certaines études récentes ont exploré la manière d'optimiser le modèle d'échantillonnage [Dal+04; Mir+04; Kum+08], mais n'ont pas inclus de critère d'échantillonnage clair afin de maximiser la qualité de la reconstruction de l'image. Bien que des méthodes comme [HK19; See+10] utilisent la limite de Cramér-Rao pour les signaux épars [BE10] comme critère d'optimalité adapté, elles sont très exigeantes en termes de calcul, ce qui limite leur utilisation à l'environnement cartésien. D'autres méthodologies dans la littérature impliquent l'empilement d'une trajectoire 2D sous-échantillonnée comme l'empilement d'étoiles [Son+04; Loo8], l'empilement de spirales [Ira+95; The+99] et l'empilement de 2D SPARKLING [Laz+20a].

Dans [Laz+20a], une approche locale de l'étincelle 3D a été proposée en concevant une trajectoire unique à l'intérieur d'un cône obtenu à partir d'une parcellation de l'espace sphérique 3D k. Ensuite, tous les cônes couvrant un plan d'élévation donné ont été remplis à l'aide de la réplication de la même trajectoire. Toutefois, cette méthode ne garantit pas un modèle d'échantillonnage localement uniforme aux limites des cônes, car le problème n'est résolu que localement.

La résolution globale du problème de l'étincelle en 3D est coûteuse en termes de calcul, avec une complexité d'ordre $O(N^2)$, où N représente le nombre d'emplacements d'échantillonnage dans le modèle. Nous avons proposé une implémentation basée sur les méthodes multipolaires rapides (FMM) [Wan+19; Fon+09a] qui nous a permis de concevoir des motifs d'échantillonnage avec jusqu'à 10^7 d'échantillons dans l'espace k, ouvrant ainsi la porte aux VDS en 3D. Nous avons comparé les implémentations multi-CPU et GPU et démontré que cette dernière est optimale pour l'imagerie 3D dans le régime d'acquisition à haute résolution (600µm isotrope). Enfin, nous avons montré que cette nouvelle optimisation pour l'étincelle 3D complète surpasse les stratégies d'empilement ou l'imagerie par projection torsadée 3D par le biais d'études rétrospectives et prospectives sur un fantôme NIST et des scanners cérébraux in vivo à 3 Tesla. Globalement, la méthode proposée permet de réduire de 2,5 à 3,75 fois le temps de balayage par rapport à l'acquisition d'imagerie parallèle GRAPPA-4 à 3 Tesla, sans compromettre la qualité de l'image.

Aborder les questions relatives aux scans prospectifs pour l'imagerie non cartésienne [Cha+22a; Cha+22c; Cha+22b; Cha+]

Après avoir étendu les trajectoires des étincelles à la 3D, un écart important a été observé dans la qualité de l'image reconstruite entre les simulations rétrospectives et les scans prospectifs réels (voir [Cha+22d, Fig. 10]) dans le cas de l'imagerie pondérée par l'T_2^*. Un tel écart dans un contexte prospectif a été identifié comme étant dû à la présence d'un fort effet de hors-résonance et d'une décroissance dont les effets sont accumulés dans le contraste T_2^*-w en raison de la longueur des Echo Time (TE) utilisées pour améliorer la contribution de la susceptibilité. Cet effet est amplifié en imagerie non cartésienne et notamment en 3D SPARKLING car ces trajectoires ont des directions de lecture arbitraires, ce qui conduit à des incohérences locales dans l'espace k. Bien que ces artefacts puissent être corrigés [Sut+03] sans nécessiter de balayage supplémentaire pour ΔB_0 map [Dav+22], de telles corrections sont coûteuses en termes de calcul.

En outre, une autre limitation des trajectoires SPARKLING est que nous utilisons des contraintes affines TE, où nous limitons la trajectoire à passer par le centre de l'espace k au moment de l'écho pour obtenir des images au contraste d'imagerie cible requis. Il en résulte un suréchantillonnage important du centre de l'espace k par rapport au critère de Nyquist, qui peut être sous-optimal car les échantillons supplémentaires peuvent être affectés à des fréquences plus élevées, ce qui permet d'obtenir de meilleures reconstructions d'images avec des détails plus fins dans les structures.

Dans ce travail, nous abordons les deux problèmes décrits ci-dessus en modifiant la fonction de coût et l'ensemble des contraintes d'SPARKLING . Avec Minimized Off Resonance Effect (MORE)-SPARKLING , nous avons introduit une pondération temporelle dans le problème d'optimisation traditionnel d'SPARKLING qui prend en compte la nature temporelle des données échantillonnées. Dans Gridding of Low Frequencies (GoLF)-SPARKLING , nous avons incorporé les caractéristiques de l'échantillonnage cartésien dans le cadre d'SPARKLING

grâce à des contraintes affines plus générales et à l'adaptation de TSD pour correspondre à ces contraintes de critères de Nyquist. Les données prospectives de l'espace k ont été acquises à 3 Tesla sur des trajectoires 20 fois plus rapides que la référence cartésienne entièrement échantillonnée. Grâce à des expériences in vivo, MORE-SPARKLING nous a permis de récupérer les pertes de signal observées sur les acquisitions originales de SPARKLING à des inhomogénéités de champ B_0 plus importantes. En outre, grâce à GoLF-SPARKLING , l'échantillonnage cartésien au centre de l'espace k a permis d'améliorer la qualité de l'image reconstruite en réduisant les artefacts.

TSD pour SPARKLING [CRC21]

L'une des limites de l'algorithme SPARKLING est la nécessité de définir un TSD en tant qu'entrée de l'algorithme. Dans nos études antérieures, nous nous sommes appuyés sur des méthodes heuristiques pour définir cette densité d'échantillonnage. Cette dernière a été paramétrée pour être radialement décroissante (sur la base de connaissances préalables issues des théories CS) et ses paramètres optimaux (décroissance, coupure) ont été recherchés sur une grille au cours d'études de reconstruction rétrospective dans lesquelles la qualité de l'image a été maximisée en fonction de trajectoires optimisées pour des densités de cible variées. Cependant, cette approche est trop coûteuse en termes de calcul. De plus, avec une densité de cible paramétrée, l'espace de recherche est trop contraint, ce qui nous empêche d'obtenir des schémas d'échantillonnage spécifiques à l'organe, au contraste de l'image ou à l'orientation. Une façon de résoudre ce problème consiste à apprendre les TSD à l'aide d'approches basées sur les données.

Dans [Kno+11], les auteurs ont proposé une approche naïve pour choisir les TSD en faisant la moyenne des spectres de puissance de plusieurs images RM dans un ensemble de données. Cette méthode aboutit à des densités d'échantillonnage qui imposent un échantillonnage plus dense dans les basses fréquences. Dans [Kno+11], les auteurs ont montré que cette approche est plus performante que le VDS standard et qu'elle reste robuste à la variabilité de l'anatomie et de l'orientation. Cependant, cette méthode se concentre uniquement sur l'ensemble des données d'IRM et ne dépend pas de la technique de reconstruction. Tous les algorithmes de reconstruction de l'IRM appliquent un a priori (comme la rareté dans le domaine des ondelettes ou du gradient de l'image). Les récents algorithmes de reconstruction par apprentissage profond (DL) [Kno+20; Muc+21b; RCS20] ont appris des a priori plus complexes basés sur l'ensemble de données spécifiques à l'organe ou au contraste. Le TSD peut être plus efficace s'il prend en compte ces prieurs et impose des échantillons plus denses dans les régions où le degré d'incertitude associé à ces prieurs pour la reconstruction est plus élevé.

Plus récemment, des méthodes comme [She+20; Bah+20a] apprennent le modèle d'échantillonnage pour l'IRM d'une manière axée sur les données tout en optimisant la qualité de l'image à l'étape de la reconstruction. Dans le cadre de l'apprentissage profond, LOUPE [Bah+20a] optimise conjointement la densité d'échantillonnage et les poids d'une architecture de réseau en U pour la reconstruction de l'image. Cependant, ces études restent limitées à l'échantillonnage cartésien.

Dans ce travail, nous utilisons la densité de cible obtenue par LOUPE comme entrée de l'algorithme SPARKLING pour générer des trajectoires non cartésiennes 2D SPARKLING. Nous réalisons des études rétrospectives et les comparons avec celles qui résultent d'autres densités telles que le spectre de puissance (log-)moyen sur l'ensemble des données de l'IRM rapide. Nous effectuons une reconstruction d'image en utilisant à la fois la technique CS et l'architecture NC-PDNet [RSC21a] nouvellement développée, qui est un réseau neuronal déroulé à densité compensée pour la reconstruction non cartésienne de l'IRM. Nous concluons que la solution proposée (LOUPE+2D SPARKLING) surpasse les autres approches VDS en termes de qualité d'image.

Apprentissage conjoint du modèle d'échantillonnage de l'espace k et du réseau de reconstruction [CRC22; CC22; CC]

L'une des limites de l'algorithme SPARKLING est la nécessité de définir un TSD comme entrée de l'algorithme. Pour cela, nous avons appris la densité d'échantillonnage à l'aide de LOUPE [Bah+20a], et l'avons utilisée comme TSD pour l'algorithme SPARKLING afin de générer des trajectoires 2D SPARKLING [CRC21]. Cependant, la grille TSD a été apprise dans le domaine cartésien, alors que la trajectoire réelle optimisée n'était pas cartésienne. Comme cela pourrait conduire à des résultats sous-optimaux, il est nécessaire d'apprendre directement des trajectoires d'échantillonnage dans l'espace k conformes au matériel ainsi qu'un réseau de reconstruction d'images dans un cadre non cartésien.

À cet égard, de nouvelles méthodes [Wei+20; Wan+21; Ved+20] ont été développées pour surmonter le besoin d'estimer un TSD, grâce à l'apprentissage conjoint direct des trajectoires d'échantillonnage de l'espace k non cartésien et de la reconstruction d'images RM d'une manière axée sur les données sur l'ensemble de données IRM rapide [Zbo+18]. Dans [Wei+20; Ved+20], les auteurs ont appris conjointement les trajectoires PILOT et les paramètres U-net en tant que modèle de reconstruction pour débruiter l'image de base produite par l'adjoint de l'opérateur Nonuniform Fast Fourier Transform (NUFFT). Toutefois, cette méthode repose sur l'auto-différenciation de l'opérateur NUFFT, qui est imprécise numériquement comme observé dans [WF23], ce qui entraîne des minima locaux sous-optimaux. Cette sous-optimalité se reflète en fait dans la forme finale des trajectoires apprises, qui ne s'écartent que légèrement de leur initialisation. Dans BJORK [Wan+21], les auteurs ont utilisé [WF23] pour obtenir une approximation jacobienne plus précise de l'opérateur NUFFT pour la rétropropagation. Les deux approches susmentionnées [Wei+20; Wan+21] ont appliqué les contraintes matérielles en ajoutant des termes de pénalité à la perte qui est minimisée pendant l'apprentissage. Bien qu'il s'agisse d'une option viable, elle nécessite le réglage d'un hyperparamètre associé à chacun de ces termes de pénalité dans la fonction de coût et ne garantit pas que les trajectoires optimisées respecteront strictement ces contraintes. En outre, ces termes de pénalité affectent les gradients globaux de la fonction de perte, ce qui entraîne une sous-optimalité des trajectoires.

Dans ce travail [CC], nous avons d'abord développé un modèle générique appelé PROjection for Jointly lEarning non-Cartesian Trajectories while Optimizing Reconstructor (PROJeCTOR). Plus précisément, nous avons introduit une méthode qui apprend les trajectoires de l'espace k d'une manière axée sur les données tout en intégrant un algorithme de descente du gradient projeté [Cha+16] afin de respecter les contraintes matérielles au cours de la phase d'apprentissage. Contrairement à BJORK, nous avons appris directement les trajectoires d'échantillonnage de l'espace k et utilisé la multirésolution [Leb+19] similaire à SPARKLING pour limiter le nombre de paramètres entraînables à chaque étape. Nous avons ensuite comparé ces résultats PROJeCTOR à PILOT [Wei+20] et BJORK [Wan+21] dans Magnetic Resonance Imaging (MRI) 2D. Dans les études rétrospectives en 2D, nos nouvelles trajectoires PROJeCTOR présentent une qualité de reconstruction d'image améliorée avec un facteur d'accélération de 20 fois sur l'ensemble de données IRM rapide avec des scores SSIM de près de 0,92-0,95 par rapport à la référence cartésienne correspondante et voient également un gain de 3-4dB en PSNR par rapport aux méthodes antérieures de pointe. Dans un cadre plus contrôlé, nous avons montré l'importance de l'étape de projection lors de l'optimisation des trajectoires dans l'espace k et avons démontré sa supériorité sur les méthodes basées sur la pénalité comme PILOT et BJORK pour mettre en œuvre les contraintes matérielles. Enfin, nous avons étendu l'algorithme à la 3D et, en comparant l'optimisation aux schémas de projection basés sur l'apprentissage, nous avons démontré que les trajectoires PROJeCTOR basées sur l'apprentissage conjoint et guidées par les données sont plus performantes que les méthodes basées sur un modèle comme SPARKLING, grâce à un gain de 2 dB en PSNR et un gain de 0,02 en SSIM.

Schéma de thèse

Chapter 1: *Introduction to Magnetic Resonance Imaging* X introduit dans MRI la source du signal mesuré et la manière dont il est localisé. Par la suite, nous discuterons de la nécessité d'accélérer le balayage Magnetic Resonance (MR) et de certaines méthodes utilisées pour accélérer les balayages en MRI cartésien traditionnel.

Chapter 2: *Compressed Sensing and Non-Cartesian MRI* X introduit la reconstruction basée sur Compressed Sensing (CS) des données MRI. En outre, nous discutons de la nécessité d'échantillonner l'espace k avec VDS pour atteindre des facteurs d'accélération plus élevés, ce qui peut être réalisé en échantillonnant l'espace k le long des courbes à l'aide d'un échantillonnage non cartésien. Pour cela, nous présentons un cadre généralisé pour l'échantillonnage de l'espace k à l'aide de trajectoires et discutons de ses contraintes. Enfin, nous présentons quelques nouvelles trajectoires non cartésiennes d'échantillonnage de l'espace k proposées dans la littérature.

Chapter 3: *Deep Learning for MRI* se concentre sur certains principes fondamentaux et le formalisme de Deep Learning (DL). Nous présentons quelques architectures DL populaires et discutons des méthodes proposées dans la littérature pour reconstruire l'image MR à partir de données sous-échantillonnées de l'espace k. En outre, nous résumons également la littérature sur les techniques basées sur l'apprentissage pour optimiser les modèles et les trajectoires d'échantillonnage de l'espace k.

Nous nous concentrons ensuite sur les principales contributions de cette thèse.Dans Chapter 4:*Optimizing full 3D SPARKLING trajectories*, nous étendons l'algorithme SPARKLING à la 3D, ce qui nous permet d'accélérer le MRI dans toutes les 3 dimensions. Cependant, de forts artefacts hors résonance ont été observés dans les études prospectives, car ces trajectoires optimisées avaient des directions de lecture arbitraires.

Pour remédier à ce problème, nous présentons dans Chapter 5:*Reducing artifacts in SPARKLING* 2 nouvelles mises à jour de l'algorithme SPARKLING : MORE et GoLF.

Comme l'algorithme SPARKLING a besoin de TSD en entrée, dans Chapter 6:*Learning sampling density for 2D SPARKLING* nous présentons 4 densités candidates différentes et montrons que la méthode LOUPE apprise en fonction des données fournit la meilleure TSD. Cependant, cette méthode implique une grille TSD, qui est apprise dans un cadre cartésien et utilisée par la suite pour générer des trajectoires non cartésiennes. Cette inadéquation dans les domaines d'apprentissage où la formation est effectuée dans un cadre cartésien et où le test est effectué dans un cadre non cartésien constitue un inconvénient majeur de cette méthode.

Ce problème est résolu par l'apprentissage direct des trajectoires de l'espace k dans Chapter 7:*Learning trajectories with reconstructor*, et nous soulignons la nécessité d'une étape de projection inspirée de SPARKLING pour éviter la sous-optimalité.

Enfin, dans Chapter 8: *Conclusions and Perspectives* nous résumons les contributions de cette thèse et discutons des orientations et perspectives futures.

Dans Appendix A:*3D SPARKLING* et Appendix B:*MORE and GoLF* nous présentons quelques résultats théoriques et expérimentaux supplémentaires pour compléter les contributions du Chapitre 4 et du Chapitre 5 respectivement.

Enfin, dans Appendix C:*Software and Open Source Contributions*, je discute de certains progiciels développés au cours de la thèse et de certaines contributions à des logiciels libres. je discute de certains logiciels développés au cours de la thèse et de certaines contributions open source.

* * *
* *
*

Publications

Patents in pipeline

- **Chaithya, G R**, G. Daval-Frérot, A. Vignaud and P. Ciuciu. "Method and apparatus for performing accelerated Magnetic Resonance Imaging with reduced off-resonance effect". Patent 1000506640/EP22305592.2. Patent Application: Europe N° 22305592.2. Apr. 2022
- **Chaithya, G R**, G. Daval-Frérot, A. Vignaud and P. Ciuciu. "Method and apparatus for performing accelerated Magnetic Resonance Imaging through gridded sampling at low frequencies". Patent application in process. Dec. 2022

Accepted articles in Peer-Reviewed Journals

- **Chaithya, G R**, P. Weiss, A. Massire, A. Vignaud and P. Ciuciu. "Optimizing full 3D SPARKLING trajectories for high-resolution Magnetic Resonance imaging". In: *IEEE Transactions on Medical Imaging* (Mar. 2022)
- **Chaithya, G R** and P. Ciuciu. "Jointly learning Non-Cartesian k-space trajectories and reconstruction networks for 2D and 3D MR imaging through projection". In: *special issue on AI in MRI: Frontiers and Applications, Bioengineering*
- Z. Ramzi, **Chaithya, G R**, J.-L. Starck and P. Ciuciu. "Density-Compensated Unrolled Networks for 2D and 3D non-Cartesian MRI Reconstruction". In: *IEEE Transactions on Medical Imaging* (Jan. 2022)
- L. El Gueddari, **Chaithya, G R**, E. Chouzenoux and P. Ciuciu. "Calibration-Less Multi-Coil Compressed Sensing Magnetic Resonance Image Reconstruction Based on OSCAR Regularization". In: *Journal of Imaging* 7.3 (2021)
- S. Farrens, A. Grigis, L. El Gueddari, Z. Ramzi, **Chaithya, G R**, S. Starck, B. Sarthou, H. Cherkaoui, P. Ciuciu and J.-L. Starck. "PySAP: Python Sparse Data Analysis Package for multidisciplinary image processing". In: *Astronomy and Computing* 32 (2020), p. 100402

Articles in pipeline to Peer-Reviewed Journals

- **Chaithya, G R**, G. Daval-Frérot, A. Massire, A. Vignaud and P. Ciuciu. "Improving SPARKLING trajectories through Minimized Off-Resonance Effects and Gridding of Low Frequencies". In: *under review MRM*

Accepted papers in Peer-reviewed Conferences

- **Chaithya, G R**, Z. Ramzi and P. Ciuciu. "Learning the sampling density in 2D SPARKLING MRI acquisition for optimized image reconstruction". In: *29th European Signal Processing Conference (EUSIPCO)*. Dublin, Ireland, Sept. 2021, pp. 960–964
- **Chaithya, G R**, Z. Ramzi and P. Ciuciu. "Hybrid learning of Non-Cartesian k-space trajectory and MR image reconstruction networks". In: *2022 IEEE 19th international symposium on biomedical imaging (ISBI)*. Kolkata, India, Mar. 2022

- K. Pooja, Z. Ramzi, **Chaithya, G R** and P. Ciuciu. "MC-PDNet: Deep Unrolled Neural Network for Multi-contrast MR Image Reconstruction from Undersampled k-space data". In: *2022 IEEE 19th international symposium on biomedical imaging (ISBI)*. Kolkata, India, Mar. 2022

Abstracts in Peer-reviewed Conferences / Papers in Workshops

- L. E. Gueddari, **Chaithya, G R**, Z. Ramzi, S. Farrens, S. Starck, A. Grigis, J.-L. Starck and P. Ciuciu. "PySAP-MRI: a Python Package for MR Image Reconstruction". In: *ISMRM workshop on Data Sampling and Image Reconstruction*. Sedona, AZ, United States, Jan. 2020
- **Chaithya, G R** and P. Ciuciu. "Benchmarking learned non-Cartesian k-space trajectories and reconstruction networks". In: *ISMRM*. 3308. London, UK, May 2022
- **Chaithya, G R**, G. Daval-Frérot, A. Massire, B. Mailhe, M. Nadar, A. Vignaud and P. Ciuciu. "MORE-SPARKLING: Non-Cartesian trajectories with Minimized Off-Resonance Effects". In: *ISMRM*. 1435. London, UK, May 2022
- Z. Amor, **Chaithya, G R**, C. Le Ster, G. Daval-Frérot, N. Boulant, F. Mauconduit, C. Mirkes, P. Ciuciu and A. Vignaud. "B_0 field distortions monitoring and correction for 3D non-Cartesian fMRI acquisitions using a field camera: Application to 3D-SPARKLING at 7T". in: *ISMRM*. 2822. London, UK, May 2022
- Z. Amor, **Chaithya, G R**, B. Daval-Frérot G. Thirion, F. Mauconduit, C. Mirkes, P. Ciuciu and A. Vignaud. "Prospects of non-Cartesian 3D-SPARKLING encoding for functional MRI: A preliminary case study for retinotopic mapping". In: *ISMRM*. 2823. London, UK, May 2022
- Z. Amor, **Chaithya, G R**, B. Daval-Frérot G. Thirion, F. Mauconduit, P. Ciuciu and A. Vignaud. "3D SPARKLING for functional MRI: A pilot study for retinotopic mapping at 7T". in: *30th proceedings of ISMRM*. Glasgow, Scotland, UK, 2022
- R. Baptista, A. Vignaud, **Chaithya, G R**, G. Daval-Frérot, F. Mauconduit, M. Naudin, M. Lapert, R. Guillevin, P. Ciuciu, C. Lerman-Rabrait and F. Boumezbeur. "Evaluation of 3D SPARKLING readout for Sodium UTE MRI at ultra-high magnetic field". In: *ISMRM*. London, UK, May 2022

Part I

Context

Chapter 1

Introduction to Magnetic Resonance Imaging

MAGNETIC resonance imaging or MRI is a popular non-invasive medical imaging technique that involves strong magnetic fields and Radio Frequency (RF) waves to produce detailed images of organs of interest. These Magnetic Resonance (MR) images can be crucial in diagnosis of a variety of disorders that occur in different body regions like brain, neck, spine, abdomen or the msucotoskeletal system. One of the strong motivations for using Magnetic Resonance Imaging (MRI) is its non-invasiveness as it does not involve any radiation exposure for the patient. Further, MRI is a versatile tool for physicians as it can be used to image different tissues in a variety of organs with a large and complementary set of weighting contrasts. Additionally, with advances in MR technology, high resolution imaging up to 0.1 mm isotropic *ex vivo* [Edl+19] and 0.2mm isotropic *in vivo* [Stu+15] becomes feasible which allows for finer biomarker delineation and quantification for diagnosis and patient follow-up, such as abnormal structural and/or functional connectivity in the brain of patients with neurodegenerative condition.

In this chapter, we discuss how an MRI system works, and the underlying physics principles used for imaging. Particularly, we describe the source of the MR signal and how it is localized in the image space. We emphasize on how the acquisition for MRI is not performed in the image domain or the pixel-space but rather in *k*-space, which ideally is the spatial frequency domain or multidimensional Fourier space. Later on, we discuss the importance of Signal-to-Noise Ratio (SNR) and how it can be improved using parallel

imaging. Finally, we outline the need to speed up the acquisition process for MRI and the approaches that have been developed to achieve this goal.

1.1 How does MRI work?

In this section we briefly cover the physics principles behind MRI which would later serve as a background for Chapter 2, where we discuss in detail the data acquisition and image reconstruction processes. This section is surely not exhaustive and to build a deeper understand we recommend the reader to refer to the following resources, which also served as an inspiration while writing this section:

- Bernstein et al. [BKZ04a] and Brown et al. [Bro+14]: classical MRI handbooks (the latter being usually known as Haacke et al. 1999);
- imaios.com/en/e-Courses/e-MRI [08]: an online course with a lot of explanatory videos;
- mriquestions.com [EB01]: an online course presented in the form of an FAQ;
- the dissertations of former PhD candidates in the same team at NeuroSpin, particularly Lazarus [Laz18] and Ramzi [Ram22].

Note that the below description of the physics of MRI is based on the *classical model* with tiny magnets that spin like tops, which are subject to macroscopic laws of electromagnetism. This view helps for easier understanding and provides a big picture which is sufficient for the purpose of this thesis. However, for a deep understanding of the actual underlying dynamics, the reader needs to rely on the *quantum mechanical model*, with spin states and discrete energy levels which go beyond the scope of this thesis.

1.1.1 Nuclear Magnetic Resonance

The key signal being measured in MRI is the Nuclear Magnetic Resonance (NMR) signal, which is a result of the interaction between the RF field and the nuclear spins in the body, particularly the spins associated with hydrogen atoms. The hydrogen atoms are present in the body in the form of water molecules, which are the main constituent of the human body. The magnetic moment associated with the nucleus of these hydrogen atoms can be viewed as tiny magnets as shown in Figure. 1.1(a). When these tiny magnets (i.e. hydrogen nuclei, Figure. 1.1(b)) are placed in a strong static magnetic field $\mathbf{B}_0$, they align themselves with the field direction in a parallel or anti-parallel manner as shown in Figure. 1.1(c). Further, each of these hydrogen nuclei precesses around the magnetic axis characterized by the Larmor frequency $\omega_0 = \gamma \mathbf{B}_0$, where γ is the gyromagnetic ratio of the hydrogen atom and $\mathbf{B}_0$ is the strength of the static magnetic field. For hydrogen nuclei $\gamma = 42.58 MHz/T$ and generally, based on the scanners, $\mathbf{B}_0$ can vary from 1.5T, 3T, up to 7T and even 11.7T [1]. This precession of the hydrogen nuclei occurs along a cone as shown in Figure. 1.1(d), and on a macroscopic scale, this leads to a net magnetization of the body $\mathbf{M}_0$ in the direction of the magnetic field.

Ideally, we would like to measure this magnetization $\mathbf{M}_0$ directly, however it is not possible as the magnetization is very small and defined along the same direction as the large applied magnetic field strength $\mathbf{B}_0$. For this reason, we rely on the resonance phenomenon RF pulse of frequency ω_1 to the body, leading to additional pulsating magnetic fields $\mathbf{B}_1$, which are perpendicular to the applied magnetic field $\mathbf{B}_0$. To induce a resonance phenomenon, the frequency of this pulse must equate the Larmor frequency, $\omega_1 = \omega_0$, i.e. the Larmor frequency in the $\mathbf{B}_0$ field in order to tip the magnetization vector $\mathbf{M}_0$ in a plane orthogonal to $\mathbf{B}_0$. After the excitation, the RF pulse is turned off and the spins enter the relaxation phase where the $\mathbf{M}_0$ vector precesses around the $\mathbf{B}_0$ field and relaxes back to the equilibrium position (Figure. 1.2(a)). During this process, orthogonal RF coils in the transverse plane are used to measure the tipped signal, called Free Induction Decay (FID) signal Figure. 1.2(c).

[1] See this press release: www.cea.fr/english/Pages/News/premieres-images-irm-iseult-2021.aspx

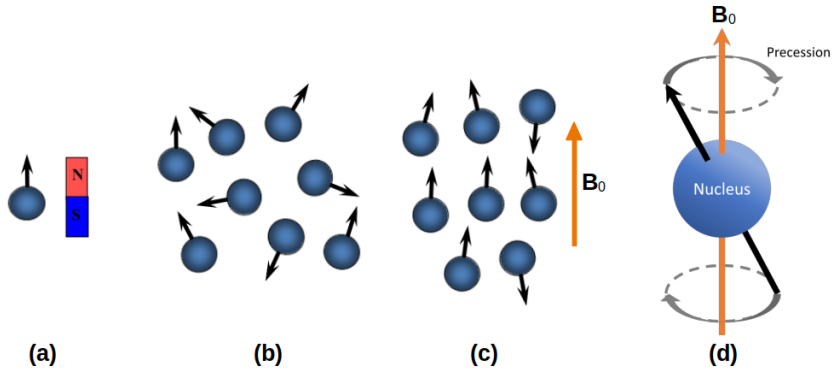


Figure 1.1 – A hydrogen atom (a) and associated magnetic field, which can be modelled as a bar magnet. (b) The bunch of hydrogen nuclei in absence of $\mathbf{B}_0$. When applying $\mathbf{B}_0$, the spin of these nuclei aligns in a parallel or anti-parallel manner to this field (c). Further, the hydrogen atom precesses around the direction of the applied field at frequency ω_0.

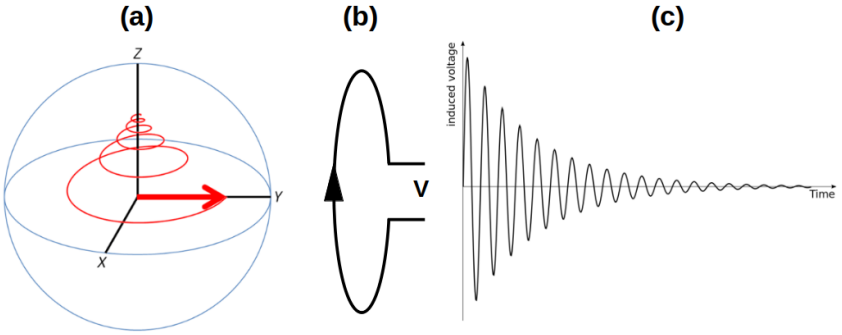


Figure 1.2 – (a) The relaxation of a spin after RF pulse is switched off. (b) The RF receiver coil used to measure the FID signal (c).

Mathematically, when the RF field is applied, the magnetization $\mathbf{M}_0$ is split into the traverse component $\mathbf{M_{xy}}$ and the longitudinal component $\mathbf{M_z}$. We write the following Bloch equations to describe the dynamics of the magnetization vector $\mathbf{M}_0$ during relaxation:

$$\frac{\mathrm{d}\boldsymbol{M_{xy}}}{\mathrm{d}t} = -\frac{\boldsymbol{M_{xy}}}{T_2} \tag{1.1}$$

$$\frac{\mathrm{d}\boldsymbol{M_z}}{\mathrm{d}t} = \frac{\boldsymbol{M}_0 - \boldsymbol{M_z}}{T_1} \tag{1.2}$$

where T_1 and T_2 are the longitudinal and transverse relaxation times respectively, which vary as a function of the tissue type. Particularly, T_1 captures the interactions of the spins with their surrounding lattice and is referred to as *spin-lattice relaxation time*, while T_2 captures the loss of phase coherence between the rotating spins in transverse plane and is called *spin-spin relaxation time*. However, in practice, the applied field $\mathbf{B}_0$ is not uniform throughout the body, leading to a faster decay of the transverse decay than expected, which is modelled using T_2^* that is linked to T_2 as follows:

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \gamma \Delta \mathbf{B}_0(\mathbf{r}) \tag{1.3}$$

where $\Delta \mathbf{B}_0(\mathbf{r})$ is the magnetic field fluctuation in space at position $\mathbf{r}$.

Solving Eq. (1.1) at a specific position $\mathbf{r} = [x, y, z]^T$ gives the following longitudinal and

transverse magnetization profiles:

$$\mathbf{M_{xy}}(t) = \mathbf{M}_0(t, \mathbf{r})\mathrm{e}^{-t/T_2} \tag{1.4}$$

$$\mathbf{M_z}(t) = \mathbf{M}_0(t, \mathbf{r})\left(1 - \mathrm{e}^{-t/T_1}\right) \tag{1.5}$$

The measured FID signal is measured in orthogonal RF coils is the projections of $\mathbf{M_{xy}}$ along x and y axes and is given by:

$$\mathbf{M_x}(t) = \mathbf{M}_0(t, \mathbf{r})\mathrm{e}^{-t/T_2}\sin\omega_0 t \tag{1.6}$$

$$\mathbf{M_y}(t) = \mathbf{M}_0(t, \mathbf{r})\mathrm{e}^{-t/T_2}\cos\omega_0 t\,. \tag{1.7}$$

1.1.2 Localizing the MR signal

Notice that the magnetization $\mathbf{M}_0(t, \mathbf{r})$ signal in Eq. (1.6) varies with spatial position $\mathbf{r}$. However, the above measured signals $\mathbf{M_x}(t)$ and $\mathbf{M_y}(t)$ at the RF receiver is a sum of all the signals in the excited volume, and hence cannot be used to localize the signal spatially. In this section, we will discuss how the acquired signal can be spatially localized through spatial encoding.

Spatial encoding using magnetic gradients

In order to distinguish the signal from different locations, the fundamental property of the spin related to their Larmor frequency ω_0 and its proportionality to the magnetic field strength $\mathbf{B}_0$ is used. This Larmor frequency is varied spatially through the application of spatially varying magnetic field gradients $\mathbf{G_x}$, $\mathbf{G_y}$ and $\mathbf{G_z}$ along each axis x, y and z respectively. These gradients result in an additional magnetic field over $\mathbf{B}_0$, that varies linearly with respect to space. For example, the gradient $\mathbf{G}_x$ along x results in the following magnetic field variation given by $\mathbf{B}_x = \mathbf{B}_0 + \mathbf{G}_x\mathbf{x}$.

In the following sections, we discuss how these gradients can be used for 2D and 3D MRI and introduce the notion of k-space. However, here for the sake of simplicity, we limit ourselves to Cartesian imaging, which consists in collecting data on a regular grid. We will introduce the concepts of non-Cartesian imaging framework later in Chapter 2 through generalization of foundations we describe below.

2D imaging

In 2D MRI, the *slice-selecting* gradient G_{slice} is applied along the z axis, which results in Larmor frequencies to vary linearly with z as $\omega_0(\mathbf{r}) = \gamma\left(\mathbf{B}_0 + G_{\text{slice}}z\right)$. Only a slice of the body or the 2D FOV along plane $z = z_0$ is selectively excited by applying a RF pulse with a frequency $\omega_0(z_0)$. Going forward, we refer to $\omega_0(z_0)$ as just ω_0 for mathematical brevity.

Later, the RF signal is turned off and a *phase-encoding* gradient G_{phase} is applied along the y axis, which causes the excited spins to precess slightly at different rates given by $\omega_0(\mathbf{r}) = \omega_0 + \gamma G_\phi y$. Once this gradient G_ϕ is turned off, the spins return to precess at $\omega_0(z)$, but now have a spatially varying phase along y axis given by:

$$\phi_y(\mathbf{r}, t) = \omega_0 t - \gamma y t_\phi G_\phi \tag{1.8}$$

where t_ϕ is the time for which the gradient G_ϕ is applied.

Finally, during the signal measurement with the RF coil, a *frequency-encoding* gradient G_{freq} is applied along the x axis, which causes the spins to precess at different rates along x axis given by $\omega_0(\mathbf{r}) = \omega_0(z) - \gamma G_{\text{freq}}x$. Due to this, the spins precess at different rates along x axis, and have varying phases along y axis. The measurement time for the signal at the receiver RF coil is called the *readout time* or Observation time (T_{Obs}). The whole timing diagram of applying the gradients and RF pulses is shown in Figure. 1.3 and is called a chronogram.

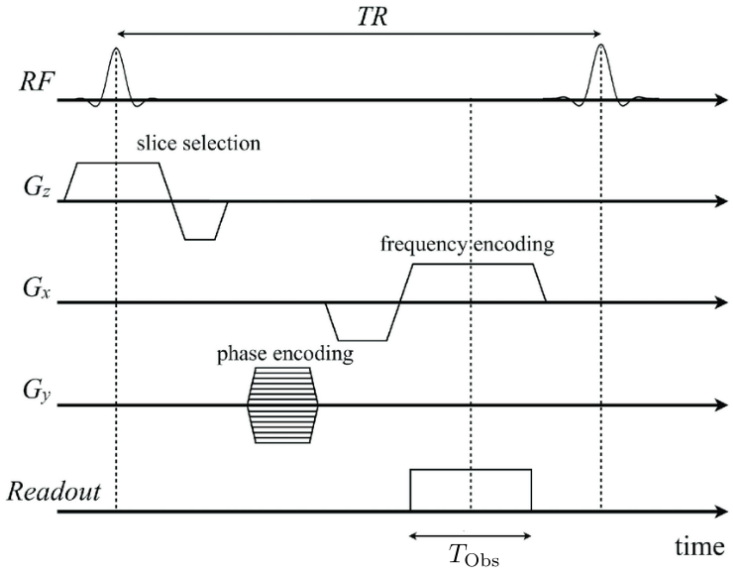


Figure 1.3 – The MR pulse sequence diagram for 2D imaging. The timing profiles of the RF pulse, G_{slice} applied along z axis (G_z), G_ϕ applied along y axis (G_y) and G_{freq} along x axis (G_x) are shown. (This figure is slightly modified version of that shown in [Pui+21])

Overall, the phase shift induced on spins at location $\mathbf{r} = [x, y]^T$ is given by (we ignore the z component as this is 2D MRI, and we only excite a slice of the body):

$$\Delta\phi(\mathbf{r}, t) = \omega_0 t - \gamma \left(y t_\phi G_\phi - x t G_{\text{freq}} \right) . \tag{1.9}$$

If $\rho(\mathbf{r})$ corresponds to the number of spins at location $\mathbf{r}$ and ignoring effect of T_2 relaxation for mathematical simplicity, the acquired signal in the RF coil is given by the sum of all the excited spins:

$$S(t) = \int_{\text{FOV}} \rho(\mathbf{r}) \mathrm{e}^{i\Delta\phi(\mathbf{r},t)} d\mathbf{r} \tag{1.10}$$

$$S(t) = \int_{\text{FOV}} \rho(x, y) \mathrm{e}^{i(\omega_0 t - \gamma(y t_\phi G_\phi - x t G_{\text{freq}}))} dx dy \tag{1.11}$$

The phase factor of $\mathrm{e}^{i\omega_0 t}$ is a modulation factor representing the Larmor precession of spins and this signal can be demodulated using a carrier frequency of ω_0, to yield:

$$s(t) = \int_{\text{FOV}} \rho(x, y) \mathrm{e}^{-ik_x(t)\cdot x - ik_y(t_\phi)\cdot y} dx dy \tag{1.12}$$

where $k_x(t) = \gamma G_{\text{freq}} t$ and $k_y(t_\phi) = \gamma G_\phi t_\phi$.

If we consider ℓ and $k_\ell \ \forall \ \ell \in \{x, y\}$ as respective Fourier conjugates, then Eq. (1.12) is a 2D Fourier transform of the $\rho(x, y)$. With this, $\rho(x, y)$ or the "MR image" can be reconstructed from the acquired signal $s(t)$ using the Inverse Fast Fourier Transform (IFFT).

Due to Eq. (1.12), we need to acquire data in the (k_x, k_y) domain, which is conventionally denoted as k-space, prior to being able to reconstruct an image in the (x, y) plan. The letter k has been used in other fields of physics to refer to spatial frequencies, which is the reason for the popularity of this convention in the MR literature.

This k-space is just the Fourier domain of the MR image being acquired during the MR acquisition process. Notice that in the time frame associated with a single RF pulse as described above, we sample a single line in k-space. After a fixed amount of time, called Repetition Time (TR), this process is repeated with a different G_ϕ to sample another line in k-space. As the MR signal decays in time, the measurement process is segmented into multiple "shots" or "readouts" and the number of readouts is denoted by N_{ro}.

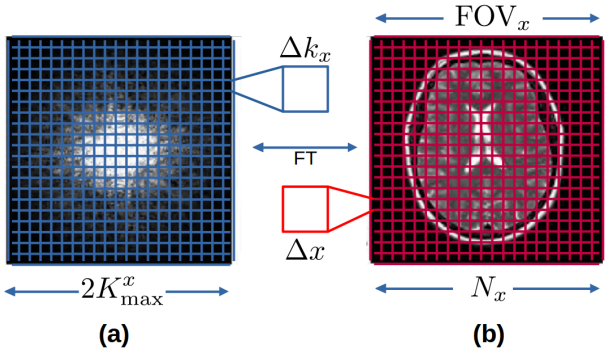


Figure 1.4 – Illustration of **(a)** k-space, and its corresponding **(b)** MR image related through Fourier transform (FT). We mark the FOV, k-space maximum $K_{\max}$ and resolution in k-space and image space.

3D imaging

The steps mentioned in the previous section cannot be carried out to reach high isotropic spatial resolution using just 2D imaging. In this setting, MRI instead consists in exciting the whole 3D FOV and acquiring data in a three-dimensional (3D) k-space.

Post excitation, a *partition-encoding* gradient G_{p} is applied along the z axis for time t_p, to encode phase information along z axis. Later, G_ϕ is applied followed by acquisition during G_{freq}. This process of chaining $G_p \rightarrow G_\phi \rightarrow G_{\text{freq}}$ with $k_z(t_p) = \gamma G_{\text{p}} t_p$, modifies the acquired signal model Eq. (1.12) into:

$$s(t) = \int_{\text{FOV}} \rho(x,y,z) \mathrm{e}^{-ik_x(t)\cdot x - ik_y(t_\phi)\cdot y - ik_z(t_p)\cdot z}\, dxdydz\,. \tag{1.13}$$

Again, during the MR acquisition process, we need to sample this entire 3D k-space to reconstruct the image using a 3D IFFT.

k-space, FOV and resolution

We now proceed to discuss the relationship between the FOV, k-space and resolution. Without loss of generality, we stick to 3D MRI, for FOV given by $\text{FOV}_x \times \text{FOV}_y \times \text{FOV}_z$. If the 3D MR volume to be reconstructed is $N_x \times N_y \times N_z$-dimensional, then the resolution in each direction is given by $\Delta\ell = \frac{\text{FOV}_\ell}{N_\ell}\ \forall\ \ell \in \{x,y,z\}$.

From Shannon-Nyquist sampling theorem, with the defined FOV, the spacing between samples in k-space must be within $\Delta k_\ell = \frac{1}{\text{FOV}_\ell}\ \forall\ \ell \in \{x,y,z\}$. Further, the maximum spatial frequency that needs to be sampled is given by $K^\ell_{\max} = \frac{1}{2\Delta\ell}$. With this, we can define the k-space as $[-K^x_{\max}, K^x_{\max}] \times [-K^y_{\max}, K^y_{\max}] \times [-K^z_{\max}, K^z_{\max}]$, with $K^\ell_{\max} = \frac{N_\ell}{2FOV_\ell}$.

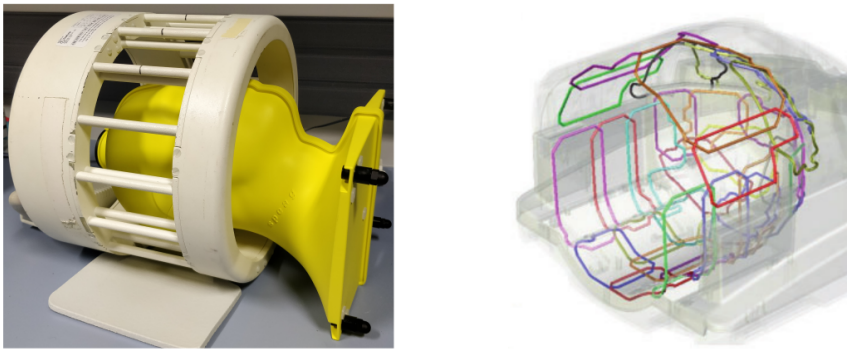
We show Figure. 1.4 to illustrate this relationship between FOV, k-space and resolution. For each readout interval, we sample a line in k-space, thus the total number of readouts $N_{ro} = N_y \times N_z$ (assuming G_{freq} is applied along x axis) and overall scan time is given by $T_{acq} = \text{TR} \times N_{ro}$.

1.2 SNR and Parallel Imaging

The acquired signal $s(t)$ in Eq. (1.13) is usually very small and in the order of *millivolts*, which results in very low SNR and poor diagnostic quality of the reconstructed MR image. The major source of noise in the measurements is due to the thermal noise at the level of the RF receiver coil which is given by

$$\sigma_{\text{noise}} \propto \sqrt{4kT_{coil}\Delta f \mathcal{R}_{eff}} \tag{1.14}$$

where k is Boltzmann's constant, T_{coil} is the temperature of the coil, Δf is the receiver bandwidth and $\mathcal{R}_{eff}$ is the effective resistance of the coil, which is a combination of coil



(a) (b)

Figure 1.5 – The MR receiver coils used for scanning the human brain. (a) A single channel birdcage coil used on a 7T MR system at NeuroSpin. (b) A 32-channel phased array coil, with individual coils are shown in color (from [Pao+15]).

resistance $\mathcal{R}_{coil}$, the electronics $\mathcal{R}_e$ and the sample being imaged $\mathcal{R}_{sample}$. Usually, $\mathcal{R}_{eff} \approx \mathcal{R}_{sample}$ which in turn is proportional to the volume of the region of body being sampled, i.e. the region in body which from where the FID signal $s(t)$ is acquired, denoted by V_{sens}. Hence, in order to boost the SNR, an improved strategy consists in choosing a smaller V_{sens}, which can be achieved by reducing the coil size and using multiple receivers in a specified configuration. These parallel imaging techniques help reach higher SNR as each individual coil has a smaller V_{sens} which reduces noise levels while still having high signal sensitivity.

1.2.1 Phased array coils

A phased array coil is a set of multiple RF coils arranged in a specific geometric pattern (see Figure. 1.5), such that they have complementary sensitivity profiles, the combination of which allow us to reconstruct a single full FOV image. Based on this geometry and the corresponding sensitivity profiles, each coil element measures the FID signal $s(t)$ for a portion of the FOV. Due to this, the V_{sens} for each coil is lower than the FOV, thereby reducing the measured noise in Eq. (1.14) and increasing the SNR while having a large effective FOV. Also, such Parallel Imaging (PI) schemes can help accelerate the scanning process whose details are given in Section. 1.4.

However, in order to obtain an optimal SNR, these phased array coils must be placed such that the noise measured across coils is largely uncorrelated, which is performed by minimizing electromagnetic interaction and coupling between the coils. Further, each receiver coil must have an individual RF receiver chain.

1.2.2 Coil compression

Most reconstruction algorithms scale linearly with the number of receiver coils Q, leading to larger reconstruction time for large coil arrays used to increase SNR. However, the redundancy in the information of the signal acquired across the multiple coils can be exploited to reduce the number of coils used for reconstruction. This is done by linearly combining the signals from different coils to reduce the coil dimensionality while having minimal loss in information, resulting in compressed effective coil signals.

One effective way to do this is by using Singular Value Decomposition (SVD), which allows us to order the compressed channels by the amount of variance, which can be a good candidate to assess information content. However, this technique could lead to high signal sensitivity and reduced image quality in the overlapping coil areas usually in the center of the measured object [Bue+07]. This issue is usually tackled through orthonormalization of the sensitivity profiles to better balance the compression error, which is efficient for imaging

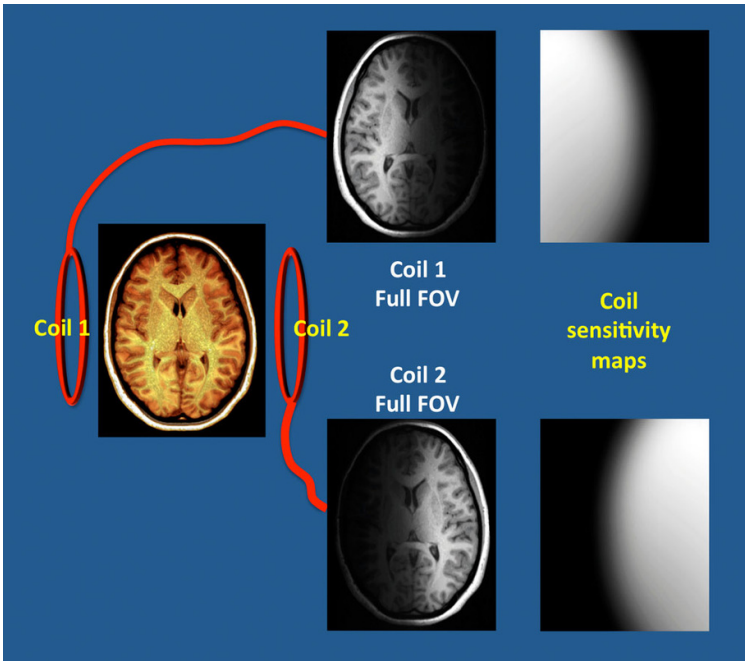


Figure 1.6 – The MR object being imaged by two coils and the corresponding per-channel images. The coil sensitivity map profiles are also shown.

and allows nearly three-fold reduction in coils without significant loss in image quality and diagnostic power.

1.2.3 Coil combination

While coil compression is helpful to reduce the number of channels to reconstruct, they cannot be used to reduce them down to a single channel as that leads to loss of information. Complex-valued image is reconstructed per compressed channel using IFFT and then combined through coil combination to form the final image. Note that coil combination algorithms compile the reconstructed images from each channel into one image for diagnosis, while coil compression methods are applied to reduce the computational load of parallel imaging.

Most coil combination techniques require the prior knowledge of the coil acquisition profiles [Roe+90] (Figure. 1.6[2]), called coil sensitivity maps $S_\ell \ \forall \ell \in \{1, 2, \dots Q\}$, which depend on the coil geometry and the object being scanned. Hence, we require a separate scan to obtain the sensitivity profiles in each exam. With the senstivity maps S_ℓ, we can recombine the MR image f from the coil specific images f_ℓ as:

$$f(\mathbf{r}) = \frac{\sum_{\ell_1=1}^{Q} \sum_{\ell_2=1}^{Q} S_{\ell_1}^H(\mathbf{r}) \Sigma_{\ell_1,\ell_2}^{-1} f_{\ell_2}(r)}{\sum_{\ell_1=1}^{Q} \sum_{\ell_2=1}^{Q} S_{\ell_1}^H(\mathbf{r}) \Sigma_{\ell_1,\ell_2}^{-1} S_{\ell_1}(\mathbf{r})} \tag{1.15}$$

where Σ_{ℓ_1,ℓ_2} is the noise correlation profiles between ℓ_1^{th} and ℓ_2^{th} coils, measured at the beginning of scan and ^{H} is the conjugate transpose operator.

The most common methods to obtain magnitude only images involves combining the images through Sum-of-Squares (SoS):

$$f(\mathbf{r}) = \sqrt{\sum_{\ell=1}^{Q} |f_\ell(\mathbf{r})|^2}\,. \tag{1.16}$$

[2] from https://mriquestions.com/senseasset.html

However, if the whole phase of the image is required, then virtual coil combination as proposed by [Par+14] can be used to establish a "virtual" coil and synchronize the phase of original or compressed coil images.

Note that all the above methods require the prior knowledge of the coil sensitivity maps, which can be obtained using a separate scan. However, there exists self-calibrated methods that can directly estimate the coil sensitivity maps from the data itself [Uec+14; El +18b]. Additionally, alternative approaches called calibrationless reconstruction techniques get rid of the need for this prior knowledge on the coil sensitivity maps and reconstruct as many images as the number of coil elements by enforcing group sparsity in the wavelet domain for instance [El +21a].

Overall, a diverse set of coil combination methods exist, and for a formal review, we invite the reader to dive into [Rob+17].

1.3 Need to speed up MR acquisition

From previous section, we see that the MR acquisition process involves sequential sampling of the lines in k-space. Due to this, achieving high resolution isotropic 3D MR imaging requires shortening scan times. For the specific case of Susceptibility Weighted Imaging (SWI), where TR is in the range of 30-40ms, for an isotropic resolution of 0.6mm for human brain with FOV of $230 \times 230 \times 124\text{mm}^3$, the scan time is nearly 50 minutes. Further, according to NHS [18], the times for MR can vary from 15 minutes to 1.5 hours. Such long scan time for a patient to stay still in the MR system, is not feasible for many clinical applications. This calls for the need to speed up the MR acquisition process for the following important reasons:

- **Patient throughput and cost**: Faster MR scans would imply a higher patient throughput, which increases the utilization of the costly MR machine.
- **Accessibility**: Some patients like people suffering from Parkinson's disease, young children and the elderly people may not stay still in the scanner for a long time. Additionally, patients with claustrophobia cannot tolerate long scan times as the confined space could cause panic attacks and anxiety.
- **Motion**: The probablity of motion increases with longer scan times, which can lead to motion artifacts, which is one of the primary sources of image quality degradation in MRI.
- **Patient comfort**: The patient needs to be present in a claustrophobia inducing MR scanner till the end of scan, which can be very uncomfortable.

For these reasons, throughout the MR literature, multiple methods have been employed to accelerate the MR acquisition process.

1.4 Cartesian sampling and acceleration

In Section. 1.1.2, we described the most basic MR acquisition process, wherein we sample the k-space sequentially along lines. This results in sampling of the k-space at specific gridded locations called Cartesian voxels, which is the most common sampling scheme in MRI, called Cartesian sampling. However, as described in Section. 1.3, this naive line by line sequential approach to acquisition leads to long scan times which is not feasible for many clinical applications.

1.4.1 Longer readouts

One common method to accelerate the acquisition process is to acquire multiple lines of the k-space in a single readout after RF pulse as done in Echo Planar Imaging (EPI) [STM91]. In EPI, a strong switched *frequency-encoding* gradient is applied simultaneously with an intermittent blip low magnitude *phase-encoding* gradient, to acquire multiple lines of the

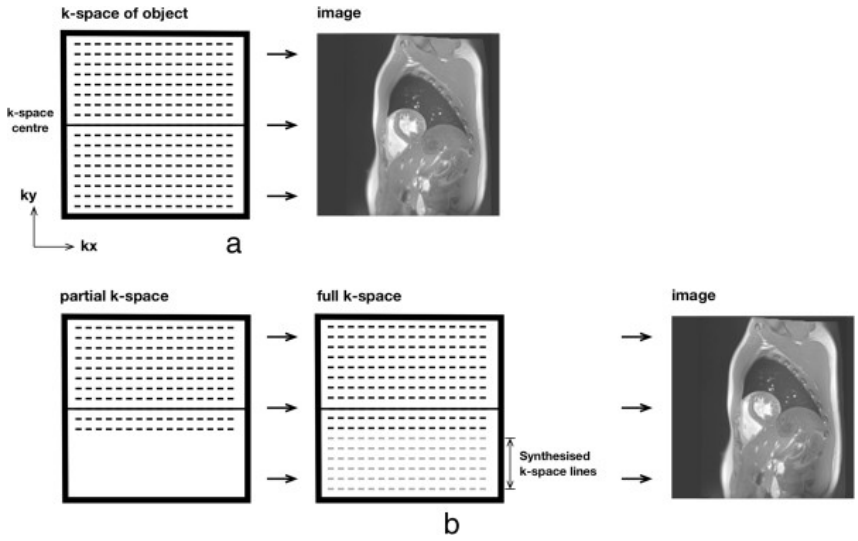


Figure 1.7 – (a) Full k-space imaging and reconstruction. Partial Fourier imaging, where missing k-space lines are synthesized from acquired data using conjugate symmetry (from [Fer+13]).

k-space in a single readout in the form of raster scans. However, this method is limited to scans with a larger time for readout, called T_{Obs}. Further, the reconstructed MR images from EPI are susceptible to geometric distortions and off-resonance artifacts dues to this longer T_{Obs}. We discuss off-resonance and its effects in detail in the next chapter. Conventionally, EPI is used in dynamic imaging applications like Functional MRI (fMRI).

1.4.2 Partial Fourier imaging

In most scenarios, as the object being imaged is a real-valued image, the corresponding Fourier domain or the k-space is redundant and possesses conjugate symmetry. Due to this, the effective number of samples required to reconstruct the image is only half of the total number of samples in the k-space, thereby reducing the scan time by theoretically one half [Fei+86], however in practice it's often less.

However, due to measurement noise, physiological motion and inhomogeneities phase errors exist and conjugate symmetry approximations are not perfect. Additionally, some imaging modalities rely on the magnetic susceptibility variations, like SWI which require the phase information of the image being reconstructed for diagnostic use. Hence, in practice, partial Fourier imaging techniques involve sampling slightly more than half of the lines in k-space typically varying from 60-75% of the total number of lines in k-space. These extra lines are later used to generate phase correction maps to more accurately predict the missing values [McG+93] (Figure. 1.7).

1.4.3 Parallel Imaging based techniques

Another popular strategy in the MRI community is to exploit the spatial sensitivity of the MR coils to accelerate the acquisition process. For doing so, during acquisition, a subset of the k-space is acquired through uniform under-sampling, where only one k-space line is acquired after skipping a fixed number of lines based on the chosen acceleration factor. This leads to strong aliasing artifacts in the k-space along the phase encoding direction (that direction along which under-sampling is implemented), which is later corrected using the spatial sensitivity of the MR coils. For this, the central portion of the k-space is acquired for calibration (used directly in [Gri+02]) or measuring a low resolution sensitivity map (used in [Pru+99]). This information from central k-space can later be utilized to fill up the missing values in the k-space across coils linearly [Gri+02]. Such correction can also be

carried out in the image space by using the coil sensitivity maps and solving a linear system of equations [Pru+99].

Both the above methods help shorten the scan duration by up to a factor of 4 while retaining a good reconstructed image quality. Additionally, his method can easily be extended in 3D and allows for accelerations in both the phase and partition encoding directions. However, using such a uniform sub-sampling pattern, the aliasing artifacts in the image usually lie along the central lines in image space, which overlap with the reconstructed images. In order to reduce artifacts and accelerate further, in Controlled Aliasing in Parallel Imaging Results in Higher Acceleration (CAIPIRINHA) [Bre+05], the under-sampling pattern along the readout directions are acquired in a staggered fashion to shift the aliasing artifacts towards the diagonals in the image space. Further extension of this method can be carried out where we undersample even in readout direction through the use of corkscrew like patterns in acquisition of k-space rather than lines (kindly refer to the next chapter for more details on such acquisition trajectories). This process is called Wave-CAIPI [Bil+15] and can be used to accelerate the scan by up to a factor of 6 to 8. More details on recent developments in parallel imaging based techniques can be found in [HFS17].

1.4.4 Towards Compressed Sensing

The above described methods help reduce the scan times, thereby have been successfully implemented commercially in MR scanners and are in active use in clinical applications. However, the above methods do not exploit the compressibility characteristics of the MR image, which is a common property of natural images in a wavelet transformed domain. Going forward, we discuss in the next chapter the sparsity or compressibility properties of MR images and how it can be exploited to further accelerate the scan times. This intrinsically relies on the concept of Variable Density Sampling (VDS), which can be more efficiently achieved using non-Cartesian sampling, where sampling of the k-space is then performed along curves off the Cartesian grid and no longer on straight lines.

* * *
* *
*

Chapter 2

Compressed Sensing and Non-Cartesian MRI

In this chapter, we discuss in depth Compressed Sensing (CS) based reconstructions and non-Cartesian sampling techniques in Magnetic Resonance Imaging (MRI). We start by describing the foundations of CS in MRI and intuitively show that sampling the k-space through Variable Density Sampling (VDS) is ideal for maximally accelerating the acquisition process, with minimal degradation in the reconstructed image quality.

Efficient way to achieve this is through non-Cartesian sampling techniques, where the k-space is sampled along curves that are not necessarily on the Cartesian grid. For this,

we generalize the notion of k-space sampling established in the previous chapter to form foundations for sampling the k-space along curves rather than lines. Later, we discuss the constraints imposed on these sampling curves due to gradient hardware, imaging contrast or physiological constraints.

We then highlight the major issue met in non-Cartesian sampling, i.e. the presence of amplified off-resonance artifacts. We discuss the causes of these artifacts, their modeling and correction strategies. Finally, we briefly introduce methods to analyze the performance of k-space sampling patterns. Particularly, we discuss the use of Point Spread Function (PSF) analysis in the absence of image datasets, which can be an accurate tool for comparing multiple sampling patterns, their performances and robustness to off-resonance effects. Additionally, image quality metrics can be used to assess the retrospective performance of the sampling pattern and corresponding reconstruction.

Throughout the chapter, our presentation will be tailored to the case of 3D MRI, while it can be of course applied to 2D MRI as well after a *slice-selective* gradient pulse.

2.1 Compressed sensing based reconstruction

CS theory has been applied and widely used in MRI [Lus+05b; LDP07] for accelerating acquisitions with minimum loss in reconstructed image quality. From the previous chapter, in an idealized scenario, M k-space data samples $\mathbf{y} \in \mathbb{C}^M$ of an image $\mathbf{x} \in \mathbb{C}^N$ with N voxels can be modeled through a Fourier operator $F_\Omega \in \mathbb{C}^{M \times N}$ as follows:

$$\mathbf{y} = F_\Omega \mathbf{x} \tag{2.1}$$

where Ω is the set of M measured k-space samples. We now need to recover an image $\hat{\mathbf{x}}$ as close as possible to $\mathbf{x}$ from the k-space data $\mathbf{y}$, under accelerated acquisition scenario with $M << N$. This problem is ill-posed with infinite number of solutions, and some prior knowledge on $\mathbf{x}$ is needed to recover a satisfactory $\hat{\mathbf{x}}$ such that $\|\hat{\mathbf{x}} - \mathbf{x}\|$ In the case of MRI like in other imaging fields, the sparsity or the compressibility of $\mathbf{x}$ in the wavelet domain plays a key role, the difference between the two notions being clarified hereafter. We proceed to detail on some fundamental concepts needed to understand CS based reconstruction.

2.1.1 Sparse representation

CS is based on the idea that the image $\mathbf{x}$ can be represented as a sparse linear combination of a set of basis functions $\Psi \in \mathbb{C}^{N \times K}$ as $\mathbf{z} = \Psi x$, where K is the number of basis functions. The coefficients $\mathbf{z}$ of this linear combination are the sparse representation of the image $\mathbf{x}$, i.e. the values of $\mathbf{z}$ are mostly zeros except for a few $s \ll K$ atoms where it has non-zero values. If such a linear operator exists, then the image $\mathbf{x}$ in the noise-free model Eq. (2.1) can be perfectly reconstructed as follows:

$$\hat{\mathbf{z}} = \underset{\mathbf{z} \in \mathbb{C}^K}{\arg\min} \|\mathbf{z}\|_0 \quad \text{such that} \quad F_\Omega \Psi^H \mathbf{z} = \mathbf{y} \tag{2.2}$$

$$\hat{\mathbf{x}} = \Psi \hat{\mathbf{z}} \tag{2.3}$$

where Ψ^H is the Hermitian transpose of Ψ. Generally, ℓ_1 relaxation is applied to the above problem as the ℓ_0 norm leads to NP-hard optimization problem [FR13].

However, in reality the acquired k-space data is noisy and also practically, the coefficients $\mathbf{z}$ are not exactly sparse but instead compressible, wich means most values being very close to zero but not exactly zero due to the presence of noise. In this more realistic scenario, the optimization problem becomes:

$$\widehat{\mathbf{z}} = \underset{\mathbf{z} \in \mathbb{C}^N}{\arg\min} \frac{1}{2} \|F_\Omega \Psi^H \mathbf{z} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{z}\|_1 \tag{2.4}$$

where λ is a regularization parameter used to balance the *data consistency* term at the left and the *regularization* term at the right. The factor of 0.5 is added to the data consistency term

to simplify the derivation of the gradient of the objective function. This way of expressing the problem as optimizing the analysis coefficients is classically known as the *synthesis* formulation. On the contrary, the *analysis* formulation is to optimize the image directly, where $\mathbf{z}$ is replaced with $\Phi\mathbf{x}$, which is more conveniently used when Ψ is overcomplete, which generally yields better reconstructed images [SF09] at the expense of longer computation times (see [EMR07; SF09; Che+18] for detailed discussions).

In the case of multi-coil acquisitions involving Q coils with sensitivity maps S_ℓ, we obtain k-space measurements $\mathbf{y}_\ell \; \forall \ell \in \{1, 2, \dots Q\}$. The reconstruction problem then becomes:

$$\widehat{\mathbf{z}} = \underset{\mathbf{z}\in\mathbb{C}^N}{\arg\min} \frac{1}{2} \sum_{\ell=1}^{Q} \|F_\Omega S_\ell \Psi^H \mathbf{z} - \mathbf{y}_\ell\|_2^2 + \lambda \|\mathbf{z}\|_1 \,. \tag{2.5}$$

The sensitivity maps S_ℓ can be estimated from the data itself using ESPIRiT [Uec+13] or low-frequency content [El +18b] methods. The above mentioned problem can also be formulated to decompose the coefficients $\mathbf{z}$ for each coil, and apply regularization jointly across cois to result in calibrationless schemes that thus get rid of estimating sensitivity maps as a first step. Typical regularization schemes that promote structured sparsity across coils are GroupLASSO and OSCAR penalizations [El +19; El +21b], the latter allowing for a sorting of coils with respect to the information they bring up in contrast to the former which assigns uniform weitghting to all coils.

2.1.2 Basis functions Ψ and regularizer

As seen in previous section, CS relies on the presence of a sparsifying transform Ψ, which is a linear operator that maps the image $\mathbf{x}$ to a sparse representation $\mathbf{z}$. In reality, as the Magnetic Resonance (MR) image is like a natural image with locally smooth regions, a large diversity of the sparsifying operators exist and have been used in the literature.

Particularly for the case of MRI, early works [LDP07] showed that discrete cosine transform and the wavelet basis (from a wide variety of wavelets like Symlet, Morlet etc.) can serve as a good sparsifying domain. We present a 2D MR image and its decomposition in the wavelet basis in Figure. 2.1. Notice that the decomposition is sparse in the detail coefficients in the wavelet basis.

More advanced sparse domains can be learned directly from the data through Dictionary Learning [RB10; RB15], which helps to build complex object and contrast specific priors. Under this framework, the dictionary Ψ is learned from the data by learning patches which can be linearly combined in a sparse manner to yield the target image of interest. Recently, more complex priors have been learned through the help of Convolutional Neural Network (CNN) and Deep Learning (DL) [AMJ18; AÖ18; Ham+19; GOW19; Sri+20] which will be discussed in the next chapter.

2.1.3 Reconstruction algorithms

Having formulated the reconstruction problem in Eq. (2.5), and choosing appropriate sparsifying transform Ψ, we now discuss the reconstruction algorithms. A broad range of convex optimization methods exist in the literature to solve the above formulated problem which can broadly be classified into the following categories:

- Proximal gradient algorithms: Iterative Soft Thresholding Algorithm (ISTA) [DDD04], Faster ISTA (FISTA) [BT09], Subband adaptative ISTA (SISTA) [BS09], Exponential Wavelet ISTA (EWISTA) [Zha+15], Proximal Optimal Gradient Method (POGM') [KF18], Fast Composite Splitting Algorithm (FCSA) [Jia+13], etc.
- Primal-dual algorithms: Alternating Direction Method of Multipliers (ADMM) [Boy+11], Primal-Dual Hybrid Gradient (PDHG) [CP11c], Condat-Vu [Con13], etc.

Extensive review of the above reconstruction algorithms particularly for the case of MRI was performed in [Fes20]. Throughout this thesis, we rely on Symlet-8 wavelet basis and

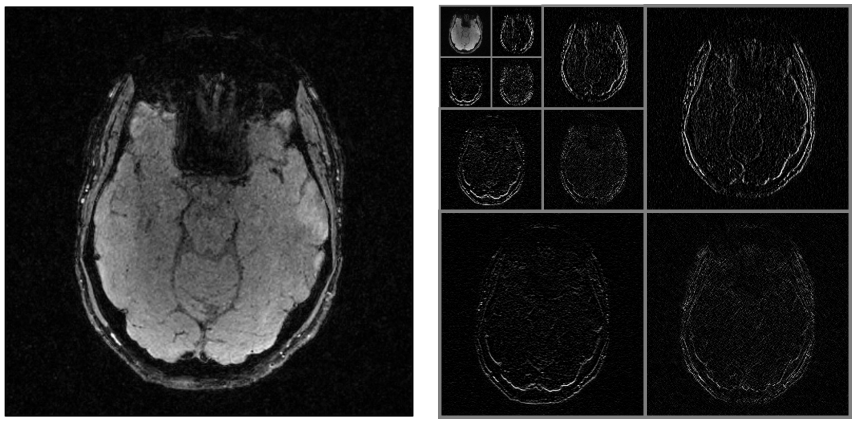


Figure 2.1 – Wavelet decomposition of an MR image (left) using the Daubechies wavelets to obtain coefficients over three scales (right). The non-sparse approximation coefficients are presented in top-left corner of this combined image, while rest of the image holds the sparse detail coefficients organized in subbands along the vertical, horizontal and diagonal axes.

FISTA [BT09] algorithm for reconstruction, which is an accelerated version of the standard proximal gradient method (also called Forward Backward). The proximal operator of a semi-continous function $\mathcal{R}$ in hilbert space is given by:

$$\text{prox}_{\mathcal{R}}(x_1) = \arg\min_{x_2} \mathcal{R}(x_2) + \frac{1}{2}\|x_1 - x_2\|_2^2 \,. \tag{2.6}$$

The proximal operator gives an improved solution for the function $\mathcal{R}$ in the vicinity of point x_1[1]. This is particularly useful when the gradient of the function $\mathcal{R}$ does not exist due to the presence of multiple subgradients.

Algorithm 1: Fast Iterative Soft Thresholding Algorithm for solving Eq. (2.5)

Inputs: N_{iter}, maximum iterations, β the step size
Output: $\hat{\mathbf{z}}$ the optimized sparse coefficients $\hat{\mathbf{x}}$ and the reconstructed image
Initializations: $\mathbf{z}^{(0)}$ an initial guess of the solution, k=0 iteration counter, $\theta^{(0)} = 1$
while $k < N_{\text{iter}}$ **do**
  // Gradient step
  $\mathbf{w}^{(k+1)} = \mathbf{z}^{(k)} - \beta \sum_{\ell=1}^{Q} \Psi S_\ell^H F_\Omega^H \left(F_\Omega S_\ell \Psi^H \mathbf{z}^{(k)} - \mathbf{y}_\ell\right)$
  // FISTA speedup using Nesterov's acceleration [Nes83]
  $\theta^{(k+1)} = \frac{1}{2}\left(1 + \sqrt{4\theta^{(k)\,2} + 1}\right)$
  // Proximal step
  $\mathbf{z}^{(k+1)} = \text{soft}_{\lambda\beta}\left(\mathbf{w}^{(k+1)} + \frac{\theta^{(k)}-1}{\theta^{(k+1)}}\left(\mathbf{w}^{(k+1)} - \mathbf{w}^{(k)}\right)\right)$
  // Iteration update
  $k = k + 1$

Having defined the proximal operator, the FISTA algorithm is described in 1, where β the step size is chosen smaller than the inverse of Lipschitz constant of the data consistency term $\frac{1}{2}\sum_{\ell=1}^{Q}\|F_\Omega S_\ell \Psi^H \mathbf{z} - \mathbf{y}_\ell\|_2^2$ in Eq. (2.5) to prevent exploding gradient. The soft thresholding operator is given by:

$$\text{soft}_\lambda(x) = \begin{cases} 0 & \text{if } |x| \le \lambda \\ x - \text{sign}(x)\lambda & \text{otherwise} \end{cases} \tag{2.7}$$

[1]We request the reader to refer to proximity-operator.net [Chi+16] to understand Proximity operators and its properties with example.

This implementation of FISTA involves speed up of ISTA [DDD04] through Nestorov acceleration [Nes83] and more improvement can be obtained through restart, enhanced momentum and greedy acceleration [LS18]. Further extensions for FISTA include faster FISTA or POGM' [Zac+19].

2.1.4 Incoherence and variable density sampling

In the above sections, we described the methods used in CS literature to reconstruct the image $\hat{\mathbf{x}}$ from undersampled k-space data $\mathbf{y}$. However, it is important to understand about the guarantees that the reconstructed image $\hat{\mathbf{x}}$ is close to the original image $\mathbf{x}$. To understand this, we define the following domains:

- **Signal space:** The space of the original image $\mathbf{x}$ is called the signal space.
- **Measurement Domain:** The space of the undersampled k-space data $\mathbf{y}$ is called the measurement space. In an ideal setting, this matches the Fourier domain of the signal $\mathbf{x}$.
- **Sparsifying domain:** The space of the sparse coefficients $\mathbf{z}$ is called the sparsifying domain. This space is characterized by transforming the image through the sparsifying transform Ψ. Typically, this is the wavelet domain.

Further, we define the *measuremnt operator* or *measurement matrix* as the linear operator that maps the sparsifying domain to the measurement space. In the case of MRI, the *measurement matrix* is given by $\mathbf{A} = F_\Omega \Psi^H \in \mathbb{C}^{M \times K}$ (we remind that M is the number of acquired k-space samples and K is the number of coefficients in the sparsifying domain).

We now try to estimate the lower bound on the probability of exact recovery of the image $\mathbf{x}$ from the undersampled k-space data $\mathbf{y}$. For this, we define the *coherence function* κ on the *measurement matrix* $\mathbf{A}$ composed with coefficients $\left((a_{i,j})_{i=1}^{M}\right)_{j=1}^{K}$ as:

$$\kappa(A) = N \times \max_{i,j}(|a_{i,j}|^2), \quad \kappa(A) \in [1, N]. \tag{2.8}$$

Then the probability of exact recovery η of the image $\mathbf{x}$ from the undersampled k-space data $\mathbf{y}$ by solving Eq. (2.5) is such that [Rau10; CP11a]:

$$M \geq C \times s \times \kappa(A) \times \log\left(\frac{N}{1-\eta}\right) \tag{2.9}$$

where s is the level of sparsity of the signal $\mathbf{x}$ in sparsifying domain, C is a constant. We note that the coherence of the *measurement matrix* $\mathbf{A}$ must be minimized to have a higher probability of recovery η.

Intuitively, we need to sample those regions in the k-space or the measurement domain, which is coherent with sparsifying domain and which cannot be reconstructed through sparsity priors. The commonly used sparsifying transform Ψ is a wavelet transform, whose detail coefficients are sparse which usually contains information on the edges or the high frequency contents of the image. However, the approximation coefficients of the wavelet transform of MR images are not sparse and need to be acquired during the acquisition process. These approximation coefficients are a low resolution version of the image of interest and holds the low frequency information of the image. Mathematically, this implies that the low frequency information is coherent between the sparsifying domain and the measurement domain. This guides the intuition that we need to sample the k-space using VDS scheme, where lower frequencies must be sampled more densely than high frequencies (please see [Puy+11; Cha+14; Adc+17; Boy+19] for mathematical details).

However, sampling patterns that obey such VDS schemes are not possible with Cartesian sampling where the acquisition is carried out using lines. For this, we need to generalize how the k-space is sampled and introduce how sampling can occur on curves, resulting in sampling locations which are off the grid, an approach referenced to as non-Cartesian sampling or imaging in the literature.

2.2 Generalizing k-space sampling

In the previous chapter, we introduced how the localization of the Nuclear Magnetic Resonance (NMR) signal was performed using spatial encoding gradients. However, for the sake of clarity, we restricted the discussion to acquiring the signal along different k-space lines. In this section, we generalize the notion of k-space sampling to include sampling along curves, which will be useful in the later sections, particularly when we introduce non-Cartesian sampling in Section. 2.3.

2.2.1 K-space sampling trajectories

Under the general setting, we apply time varying gradient fields $G_x(t)$, $G_y(t)$ and $G_z(t)$ along the x, y and z axes respectively during the acquisition of the signal. Note that we do not apply any gradients in between Radio Frequency (RF) pulse and the signal readout.

Due to this, the spins accumulate the following time varying phase difference during the acquisition time:

$$\Delta\phi(x,y,z,t) = \omega_0 t - \gamma\left(x\int_0^t G_x(\tau)d\tau + y\int_0^t G_y(\tau)d\tau + z\int_0^t G_z(\tau)d\tau\right) \tag{2.10}$$

the corresponding demodulated signal is given by:

$$s(t) = \int_{\text{FOV}} \rho(x,y,z)\mathrm{e}^{-ik_x(t)\cdot x - ik_y(t)\cdot y - ik_z(t)\cdot z}\,dxdydz \tag{2.11}$$

$$k_\ell(t) = \gamma\int_0^t G_\ell(\tau)d\tau, \quad \forall\ \ell \in \{x,y,z\} \tag{2.12}$$

Throughout the readout time, also called Observation time (T_{Obs}), this signal is sampled at discrete time rate by Analog to Digital Converter (ADC), at a dwell time pace, denoted by δt. Thus, the number of k-space samples acquired during the readout time T_{Obs} is given by $\left\lfloor \frac{T_{\text{obs}}}{\delta t} \right\rfloor$. Typically, the dwell time varies from 1 to 10 μs, depending on the Signal-to-Noise Ratio (SNR) requirements and the hardware constraints from ADC.

During each readout, the applied gradients $G_\ell(t), \forall\ \ell \in \{x,y,z\}$ control the spatial location where the k-space signal is acquired during T_{Obs}. This sampling location varies with time and is given by $\mathbf{k}(t) = [k_x(t), k_y(t), k_z(t)]^T$, which is called k-space sampling trajectory or a "shot". Different time varying gradient profiles are applied along each axis for each readout, which results in different sampling trajectories or curves, that each sample a different region in the k-space. We denote N_c as the number of shots or readouts involved in the acquisition of the field-of-view (FOV). Thus, the k-space sampling trajectories are composed of $\mathbf{k}_i(t) = (k_{i,x}(t), k_{i,y}(t), k_{i,z}(t))$, which in turn are controlled by magnetic field gradients $\mathbf{G}_i(t) = (G_{i,x}(t), G_{i,y}(t), G_{i,z}(t))$, where $i \in \{1,2,\ldots,N_c\}$. A sample k-space trajectory and its corresponding gradients are shown in Figure. 2.2.

Typically, during the acquisition process, the overall k-space sampling pattern and its corresponding trajectories are chosen, whose derivatives give the gradient profiles to be played by scanner as (see Eq. (2.12)):

$$G_{i,\ell} = \frac{1}{\gamma}\frac{dk_{i,\ell}(t)}{dt}, \quad \forall\ \ell \in \{x,y,z\}\,. \tag{2.13}$$

2.2.2 Constraints on k-space sampling trajectories

Note that these k-space sampling trajectories are defined based on the time varying gradient profiles applied during the acquisition by the gradient hardware of the scanner. Due to this, the k-space sampling trajectories are constrained by the hardware limits (see Figure. 2.3). Further, additional constraints are present due to physiological constraints and image contrast requirements. We detail each of them and present a mathematical constraint set for the k-space sampling trajectories in the following subsections. A detailed mathematical treatment for these constraints has been introduced in. [Cha+16].

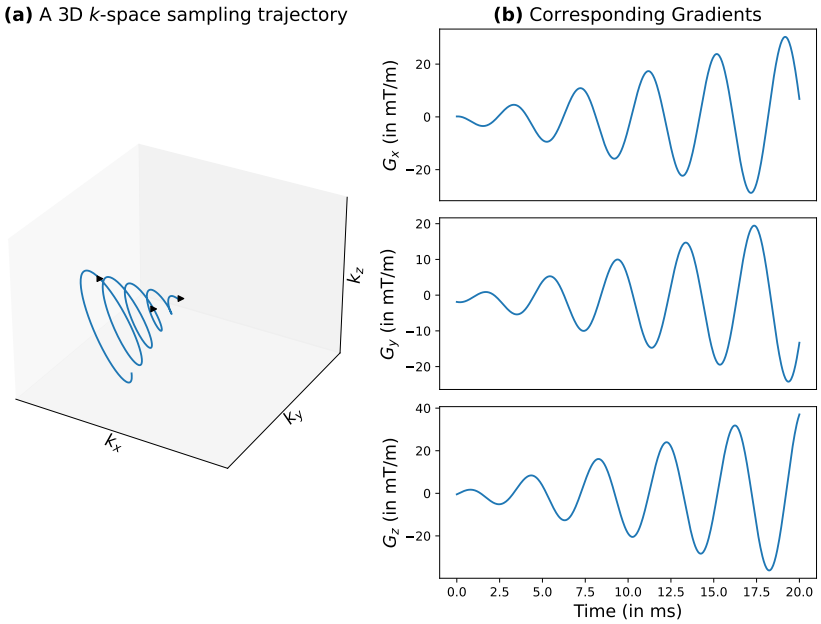


Figure 2.2 – **(a)** A single 3D k-space sampling trajectory and **(b)** its corresponding gradients in x, y and z directions obtained with Eq. (2.13), which is played during readout of time period T_{Obs}.

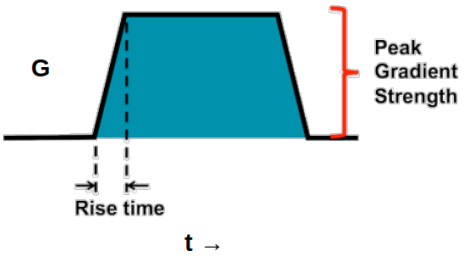


Figure 2.3 – Illustration of source of gradient constraints. The gradient amplitude is limited by the peak gradient strength $G_{\max}$. Also, the rise time results in limiting the maximum slew rate $S_{\max}$.

Gradient constraints

The gradient profiles ($G_x(t)$, $G_y(t)$ and $G_z(t)$) are played by gradient coils driven by electrical currents that are pulsed to give a trapezoidal waveform (Figure. 2.3). This underlying hardware that produces such waveforms is limited by:

- **Gradient raster time**: While the waveforms are continuous, they can be played by the hardware at a discretized time period, called gradient raster time Δt. Typically, the values of this on scanners vary from 4µs (on GE MR750 3T) to 10µs (on Siemens Magnetom Prisma$^{\text{FIT}}$ 3T). Due to this, the k-space trajectory curves are also discretized at every Δt, giving a discrete set of sample points along curve: $(\mathbf{k}_i[n])_{n=0}^{N_s-1}$, where $N_s = \left\lfloor \frac{T_{\text{obs}}}{\Delta t} \right\rfloor$ is the number of samples per shot. In Figure. 2.4, we show an example of continuous sampling trajectory in a normalized k-space (red) and corresponding gradient profile, and its discretized version with $\Delta t = 10\mu s$. Notice that intermediate acquired samples at ADC are a linearly interpolated version of this discrete trajectory.
- **Maximum gradient amplitude**: The trapezoidal current waveforms are limited by

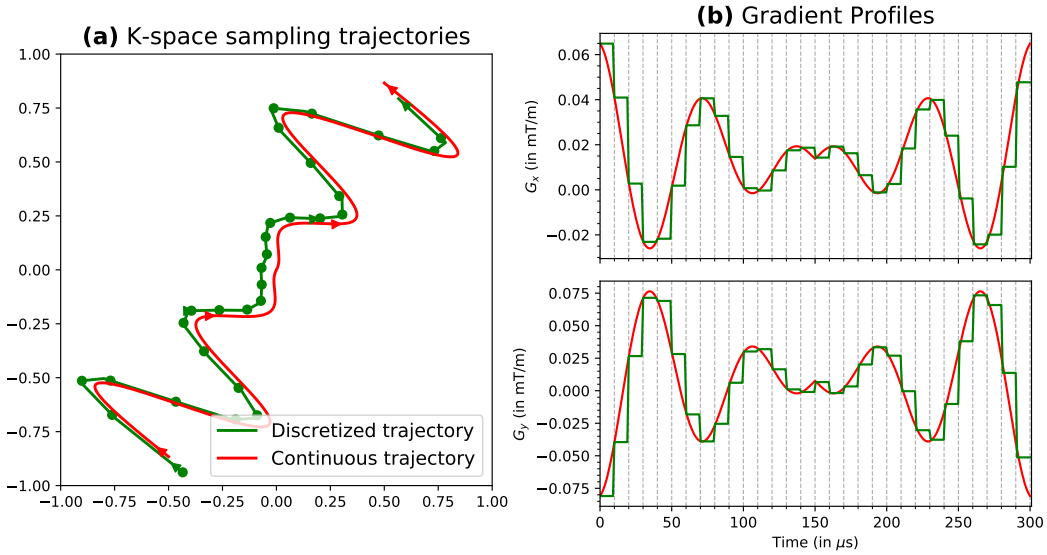


Figure 2.4 – **(a)** An example of continuous sampling trajectory in a normalized k-space (red) and corresponding gradient profile. The discretized gradient profile with a Gradient raster time $\Delta t = 10\mu s$ is shown in green and its corresponding discretized trajectory is also shown in **(a).**

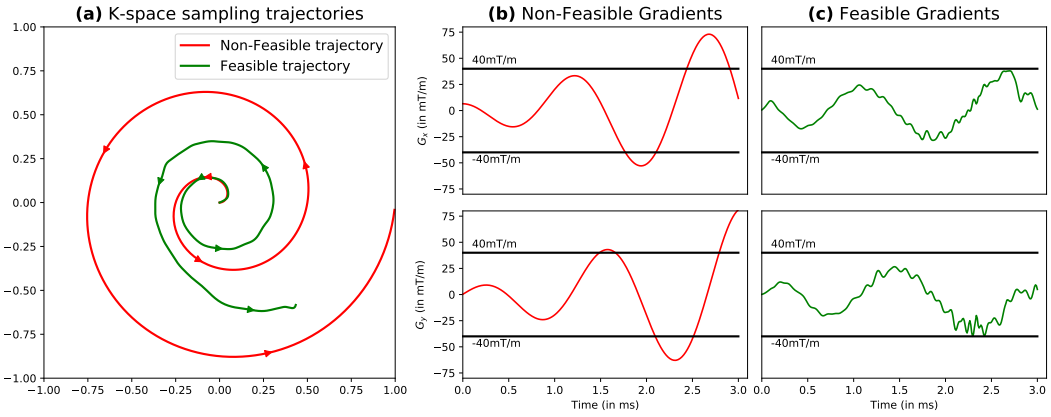


Figure 2.5 – **(a)** An example of non-feasible k-space sampling trajectory (red) which violates the maximum gradient $G_{\max} = 40mT/m$ constraint, and the closest feasible k-space trajectory (green). The corresponding gradient profiles G_x and G_y are shown in **(b)-(c)**. Also, we highlight the maximum feasible gradient amplitude $G_{\max} = 40mT/m$ that can be played by the gradient hardware with a solid black line.

maximum gradient amplitude $G_{\max}$ measured typically in mT/m. Usually, $G_{\max}$ varies from 30-45mT/m on high field scanners and 15-25mT/m on low field scanners. As a result of this constraint, the maximum playable gradient is limited which thereby limits the maximal speed at which the trajectory traverses the k-space:

$$|G_{i,\ell}[n]| = \frac{|\mathbf{k}_{i,\ell}[n] - \mathbf{k}_{i,\ell}[n-1]|}{\gamma \Delta t} \leq G_{\max}, \quad \begin{array}{c} \forall\ \ell \in \{x, y, z\}, \\ \forall\ i \in \{1, 2, \ldots, N_c\}, \\ \forall\ n \in \{0, 1, \ldots, N_s - 1\} \end{array} \tag{2.14}$$

As an example, we show in Figure. 2.5 an example of non-feasible k-space sampling trajectory (red) which violates the maximum gradient $G_{\max} = 40mT/m$ constraint, and the closest feasible k-space trajectory (green).

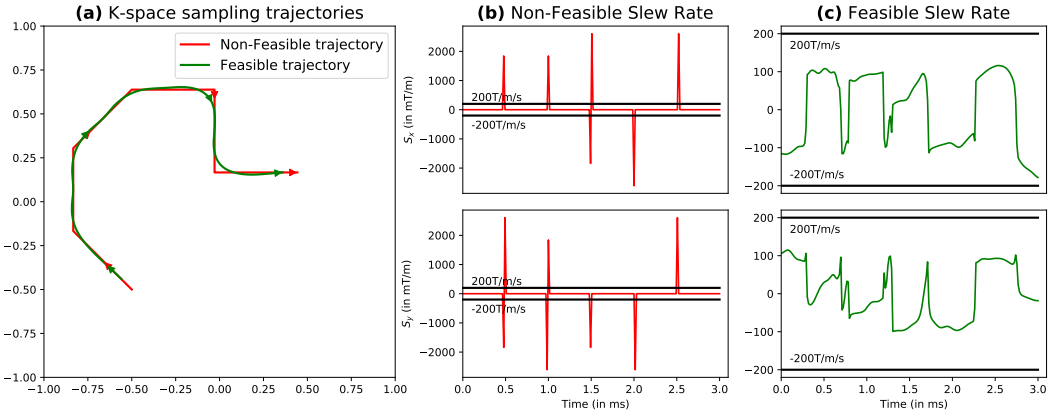


Figure 2.6 – **(a)** An example of non-feasible k-space sampling trajectory (red) which violates the maximum slew rate $S_{\max} = 200T/m/s$ constraint, and the closest feasible k-space trajectory (green). The corresponding slew rate profiles S_x and S_y are shown in **(b)-(c)**. Also, we highlight the maximum feasible gradient amplitude $S_{\max} = 200T/m/s$ that can be played by the gradient hardware with a solid black line.

- **Maximum slew rate**: The gradient hardware takes a minimum time called *rise time* t_r to change the gradient strength. This minimum rise time constraint is measured as maximum achievable slew rate $S_{\max} = G_{\max}/t_r$. Typically, the value of $S_{\max}$ varies from 150-200mT/m/s on high field scanners and 50-100mT/m/s on low field scanners. This constraint limits the rate at which the gradient can change its strength, which in turn limits the rate at which the k-space trajectory can change its direction, or its acceleration:

$$\frac{|G_{i,\ell}[n] - G_{i,\ell}[n-1]|}{\Delta t} = \frac{|\mathbf{k}_{i,\ell}[n+1] - 2\mathbf{k}_{i,\ell}[n] + \mathbf{k}_{i,\ell}[n-1]|}{\Delta t^2} \leq S_{\max} \tag{2.15}$$

  As an example, we show in Figure. 2.6 an example of non-feasible k-space sampling trajectory (red) which violates the maximum slew rate $S_{\max} = 200T/m/s$ constraint, and the closest feasible k-space trajectory (green).

Constraints for stable contrast

Note that, the underlying received signal equation (Eq. (2.11)) does not take the effect of T_2^* relaxation into account. Taking it into account, we have a decay of the magnetization of spins which is accumulated at the received signal as follows:

$$s(t) = \int_{\text{FOV}} \rho(\mathbf{r}) \mathrm{e}^{-\alpha_{\mathbf{r}} t} e^{-i(\mathbf{k}_i(t)\cdot\boldsymbol{r})}\, d\boldsymbol{r} \tag{2.16}$$

where $\alpha_{\mathbf{r}} = \frac{1}{T_2^*}$ is the spatially varying decay rate of the signal at position $\mathbf{r} = [x, y, z]^T$. In order to have a stable imaging contrast, the low frequency content of the MR image should be sampled at the same time, which usually corresponds to the Echo Time (TE) of the sequence. This imposes an affine constraint on the k-space trajectory where the trajectories are constrained to pass through the center of k-space, i.e. $[0, 0, 0]^T$ at TE during every shot:

$$\mathbf{k}_{i,\ell}[n_{\text{TE}}] = 0. \quad \forall\ \ell \in \{x, y, z\}, \forall\ i \in \{1, 2, \ldots, N_c\} \tag{2.17}$$

where $n_{\text{TE}} = \left\lfloor \frac{T_{\text{TE}}}{\Delta t} \right\rfloor$ is the index of the sample point at TE.

Physiological constraints

The varying gradient profiles applied during the MR acquisition leads to electrical potentials, which could induce excitation of nerves in the extremities, called as Peripheral Nerve

Stimulation (PNS) [SBN00]. This can be perceived by the patient as a tingling sensation, which can be uncomfortable. The levels of PNS must be controlled and limited under safe levels to prevent risking the heart.

While models exist to predict the PNS levels given the gradient profiles [HG00], this relationship is usually complex, and in practice lowering the $S_{\max}$ of the trajectory would also result in reduced PNS.

2.3 Non-Cartesian Sampling

Having generalized the k-space data acquisition in the previous section, we now discuss different sampling schemes to collect k-space data using such generalized constrained k-space curves. One of the crucial reasons to use non-Cartesian sampling is to increase efficiency of k-space coverage, allowing us to sample broader k-space regions in a shorter time. Further, sampling along curves adds a degree of freedom to the sampling scheme which can help in achieving a prescribed Target Sampling Density (TSD), which helps achieve a required VDS as recommended by CS literature [Puy+11; Cha+14; Adc+17; Boy+19]. Additionally, non-Cartesian sampling can also be used to achieve robust-to-motion sampling schemes, where the central low frequencies in k-space are sampled repetitively as performed in PROPELLER/BLADE [Hir+08; Pip99] (Figure. 2.7(c)) and Rosettes [Nol97] (Figure. 2.7(d)). Finally non-Cartesian sampling patterns allow for acquiring k-space samples with minimal TE which is useful for Magnetic Resonance Spectroscopy Imaging (MRSI) and Ultrashort Echo Time (UTE) imaging.

In this section, the focus will be to highlight some popular non-Cartesian sampling patterns, their advantages and applications. Finally, we provide a brief note on how such non-Cartesian k-space data is reconstructed to form an MR image.

2.3.1 Trajectories

While our primary focus here will be on 3D non-Cartesian k-space sampling trajectories, we still discuss some popular 2D trajectories which could be extended to 3D through stacking along slice direction or rotations. However, full 3D k-space trajectories that efficiently accelerate even in the third dimension outperform such stacked and rotated trajectories.

Radial sampling [Lau73] (Figure. 2.7(a)) is a popular sampling scheme in MR imaging, where the k-space trajectory is sampled along radial lines from the center of k-space (i.e. *center out*). This scheme was introduced even before the advent of CS theory and *parallel-imaging*. They have gained popularity in MR imaging as they naturally enforce a VDS, with oversampling at the center of k-space. Further, repeated sampling of the center of k-space allows for robust-to-motion imaging. While 3D versions of the trajectories exist (i.e. Koosh-Balls [Lar+08]), extensions to can also be achieved through stacking, resulting in stack-of-stars, which is known to be efficient for dynamic imaging like Cardiac MRI.

Spiral trajectories (Figure. 2.7(b)) are used to accelerate imaging as they result in images which are robust to motion, flow and aliasing artifacts. These trajectories take full advantage of the gradient hardware where the trajectories are limited by slew rate $S_{\max}$ in the center of k-space and maximum gradient limited at the edges of k-space [Del+10]. This is controlled through parameter choices [Glo99; KAS03; Lee+03] which remain restrained by this design on the $G_{\max}$ and $S_{\max}$ constraints.

Three-dimensional extensions of the spiral trajectories are achieved in the form of stacks [Ira+95] or cones [GHN06]. Other extensions include hybrid radial-cones [Joh17] where a single cone is distributed along the 3D k-space through rotation. More recently, Fermat Looped ORthogonal Encoded Trajectories (FLORET)[Pip+11a] (Figure. 2.7(e)) were introduced to be more efficient than conventional stack-of-cones with the same SNR efficient but half the scan time. As radial trajectories are inefficient in higher frequencies (as they leave out large gaps between radial lines), they are combined with trajectories to get TWisting Radial Lines (TWIRL) [JNM92] in 2D and Twisted Projection Imaging (TPI) [Boa+97] in 3D, which can be used to image irregular flow or in MRSI imaging.

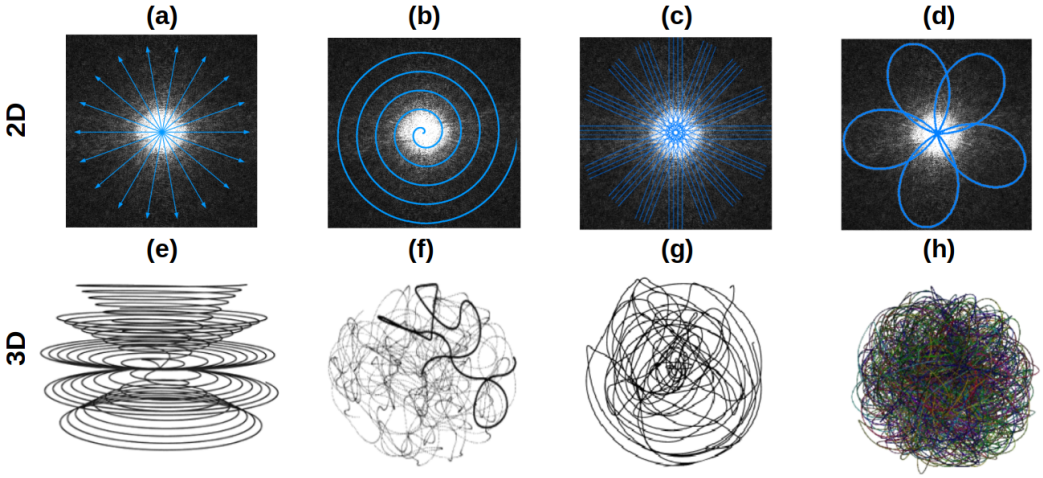


Figure 2.7 – Some 2D (top row) and 3D (bottom row) non-Cartesian k-space sampling trajectories proposed in literature. **(a)** Radial [Lau73] **(b)** Spiral [Ahn+86b] **(c)** PROPELLER [Hir+08; Pip99] **(d)** Rosettes [Nol97] **(e)** FLORET [Pip+11a] **(f)** Genetic [Dal+04] **(g)** Missile [Mir+04]**(h)** Durga [Kum+08]

In recent years, there has been a shift to design non-Cartesian trajectories through optimization that result in more efficient non-parametric sampling patterns. The optimization criteria could be either heuristic [Dal+04; Mir+04; Kum+08] (Figure. 2.7(f-h)) or chosen within a set of pseudo-random or well defined trajectories [See+10; RB11; Liu+12]. Further, the Spreading Projection Algorithm for Rapid K-space sampLING (SPARKLING) algorithm [Boy+16; Cha+17; Laz+19] has been shown to automatically generate optimized non-Cartesian sampling patterns compatible with MR hardware constraints on maximum gradient amplitude and slew rate. SPARKLING optimally samples the k-space (see [Puy+11; Cha+14]) with a controlled distribution of samples (e.g., variable density) and a locally uniform k-space coverage.

2.3.2 Reconstruction

As discussed earlier, in non-Cartesian sampling the measured k-space data is not sampled on a regular grid, and hence the reconstruction of the image from the k-space data is not straightforward. Further, accelerated schemes under-sample some k-space region, particularly the periphery when VDS is implemented. In this section, we will briefly discuss the reconstruction of non-Cartesian data to form an image.

Non-uniform fast Fourier transform

Throughout the reconstruction process, the Nonuniform Fast Fourier Transform (NUFFT) operator plays an important role as it helps us to map the k-space data which does not belong on a grid to the gridded image domain, and vice versa. This mapping involves gridding the k-space data followed by Inverse Fast Fourier Transform (IFFT). Conversely, we can also simulate the acquired non-Cartesian k-space data from an image by interpolating the Fourier domain of the image to trajectory locations.

This mandatory step of gridding the k-space data (Figure. 2.8) can be performed in two alternative ways (see [Pau] for a detailed discussion):

- **Grid-driven methods:** This is a naive approach where the value of k-space data at each grid point is computed by interpolating it from the off-the-grid measurements. These measurements are usually a subset of the k-space data at locations within a maximum parametrized from the grid point. However, under this method some data may be skipped as they are outside this parametrized radius. To prevent this, the radius

(a) Grid-driven. (b) Data-driven .

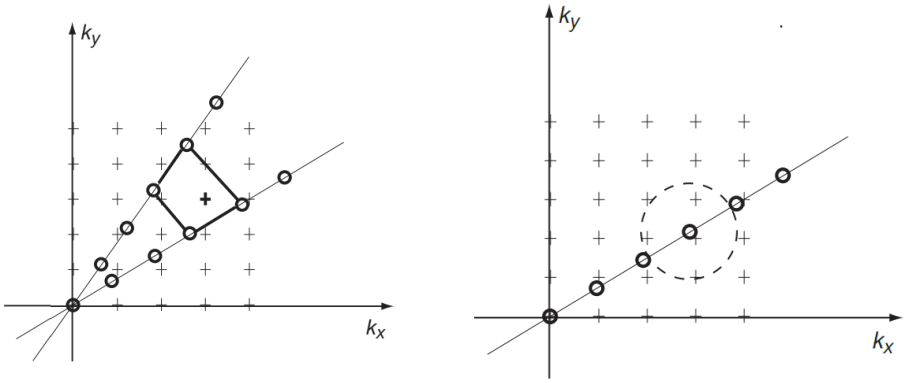


Figure 2.8 – The two approaches for gridding step in the NUFFT are represented: the **grid-driven** methods (left) compute the on-the-grid values (+) by interpolating the off-the-grid acquired samples (∘), and **data-driven** techniques (right) apply kernels to each off-the-grid sample to accumulate information over the gridded voxels.

is chosen such that there is a wide overlap between these grid point neighborhoods, which can potentially lead to inaccuracies. Other way to tackle this is to upsample the grid to a higher resolution, which can become computationally expensive.

- **Data-driven methods:** This method works in an opposite way and distributes the contributions of each k-space measurement onto the neighborhood of the grid points. Similar to grid-driven approach, upscaling the Cartesian grid can help to apply more precise approximations, however this does not increase the computational cost drastically as number of data points remain the same. A gridding kernel is chosen to interpolate the data onto the Cartesian neighborhood, and typically Kaiser-Bessel kernel is used [Jac+91; ST95; Ras+99].

  However, as this method adds the sample contributions, densely acquired data points are over-represented and there is a need to compensate for this through Density Compensation (DCp). This compensation is typically done by weighing the k-space data inversely with respect to the density of samples in the specific locations where the data was acquired. One popular scheme [PM99] to estimate this density compensator vector $\mathbf{w} = [w_m]_{m=1}^{M}$, where M is the number of k-space samples is through the following iteration having initialized the weights with $\forall m, w_m^{(1)} = 1$:

$$\mathbf{w}^{(i+1)} = \frac{\mathbf{w}^{(i)}}{\mathbf{w}^{(i)} \otimes \Psi_{\text{kb}}} \tag{2.18}$$

  with Ψ_{kb} the chosen kernel (usually Kaiser-Bessel) and $\otimes$ the convolution operator.

Over the years, there have been many implementations of the NUFFT operator that are faster, optimized and more accurate [FS03; GL04]. Many practical implementations exist for the NUFFT operator varying from CPU implementations [GL04; KKP09; Vai+23; BMK18], to GPU [Lin18; Kno+14; Shi+21] and extensions to tensor-centeric frameworks like TensorFlow [Mon22; RC23] and PyTorch [Muc+20]. Some of these libraries [Kno+14; Mon22; RC23] also have inbuilt routines to estimate the density compensator vector $\mathbf{w}$.

Algorithms

Most reconstruction algorithms for accelerated non-Cartesian MRI are direct extensions of the algorithms described in Section. 2.1, but the Fourier transform operator is replaced by the NUFFT operator.

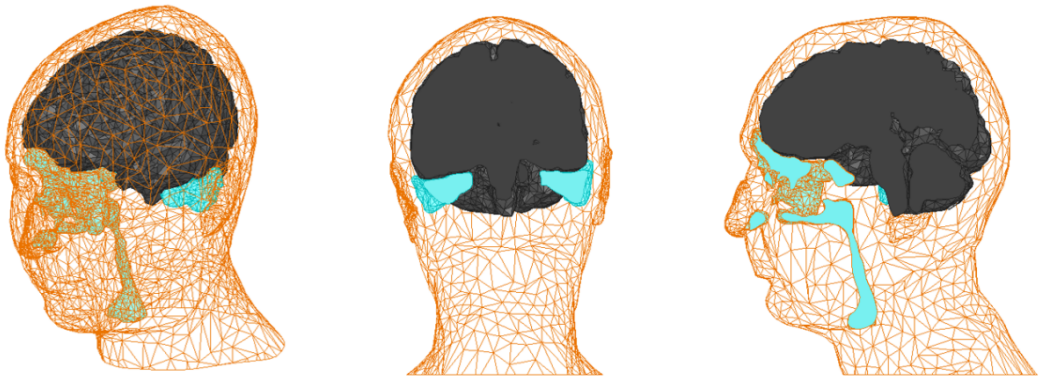


Figure 2.9 – The water/air interfaces (blue) in the brain (black), such as the ones in the ears (middle) or near the bucco-nasal cavities (right). From [Pin21].

Further, to speed up most of the optimization schemes, the adjoint operator F_Ω^H operating on the k-space data y_ℓ is pre-conditioned with the density compensator vector $\mathbf{w}$ in the iteration step in 1, as shown below:

$$\mathbf{z}^{(k+1)} = \text{soft}_{\lambda\tau}\left(\mathbf{z}^{(k)} - \tau \sum_{\ell=1}^{L} \mathbf{\Psi}\mathbf{S}_\ell^* F_\Omega^H D\left(F_\Omega \mathbf{S}_\ell \mathbf{\Psi}^* \mathbf{z}^{(k)} - \mathbf{y}_\ell\right)\right)$$

where all notations are the same as above and D is the diagonal matrix with diagonal elements as $\mathbf{w}$.

2.4 Off-resonance effects

One popular issue with non-Cartesian imaging, is the amplification of off-resonance artifacts. In this section, we briefly introduce the sources of these off-resonance effects and how they can lead to artifacts in the reconstructed MR images. We also discuss some methods in literature to mitigate these artifacts. This section will serve as a good introduction for the Chapter, where we try to tackle these artifacts in non-Cartesian imaging during trajectory design.

2.4.1 Sources and modelling

The applied B_0 magnetic field by the scanner is customized to be as homogenous spatially as possible through shimming, so that all the corresponding spins have the same expected Larmor frequency. However, air-tissue interfaces exists inside the human body particularly near ear-canals or in vicinity of bucco-nasal region, as illustrated in Figure. 2.9. These interfaces have differences in magnetic susceptibility leading to perturbations in B_0 field which leads to so-called off-resonance artifacts. The latter can considerably degrade the image quality during data acquisition.

These perturbations in B_0 field can be modeled as a spatially varying magnetic field $\Delta B_0(\mathbf{r})$ which lead to spatially varying changes in Larmor frequencies $\omega_\mathbf{r}$:

$$\Delta B_0 = \Delta\chi \times B_0 \tag{2.19}$$

$$\omega(r,t) = \gamma(B_0 + G(t)\cdot r + \Delta B_0(r))\,. \tag{2.20}$$

where $\Delta\chi$ is spatially varying profile of magnetic susceptibility variations.

The generalized signal equation with T_2^* decay, i.e. Eq. (2.16), can be extended by taking the spatially varying field perturbations $\omega_\mathbf{r}$ into account as follows:

$$s(t) = \int_{\text{FOV}} \rho(\mathbf{r}) \mathrm{e}^{-\alpha_\mathbf{r} t - i\omega_\mathbf{r} t} e^{-i(\mathbf{k}_i(t)\cdot\boldsymbol{r})}\, .d\boldsymbol{r} \tag{2.21}$$

The additional term $\mathrm{e}^{-\alpha_{\mathbf{r}}t - i\omega_{\mathbf{r}}t}$ in the signal equation is both spatially and temporally varying, which results in diverging from the conventional Fourier model of the acquired signal.

2.4.2 Estimating a ΔB_0 field map

In the absence of these effects, the MR image x can be reconstructed from the acquired signal s as:

$$x(\mathbf{r}) = \sum_{i=1}^{N_c} \int_{T_{obs}} s(t) e^{i\mathbf{k}_i(t)\cdot\mathbf{r}} \mathrm{d}t \,. \tag{2.22}$$

This corresponds to the adjoint operator of NUFFT operator. However, in the presence of off-resonance effects, the signal equation Eq. (2.21) yields a more complicated signal-to-image relation:

$$\widehat{x}(\mathbf{r}) = \sum_{i=1}^{N_c} \int_{T_{obs}} s(t) e^{i(\mathbf{k}_i(t)\cdot\mathbf{r} + \omega_{\mathbf{r}}t)} \mathrm{d}t \,. \tag{2.23}$$

The additional term $\omega_{\mathbf{r}}$ needs knowledge of ΔB_0 field map to be corrected.

Most common methods to estimate this maps involve measurement of the observed phase shift $\phi_{\Delta B_0}(\mathbf{r})$ of the MR image x at position $\mathbf{r}$. As an approximation, this observed phase at TE is related through a linear relationship [GJ16; Rob+11]:

$$\phi_{\Delta B_0}(\mathbf{r}, TE) = TE \times \omega_{\mathbf{r}} = TE \times \gamma \Delta B_0(r) \,. \tag{2.24}$$

With this, a ΔB_0 map can be estimated by performing acquisitions with multiple echo times [GJ16], and consider other phase contributions as constant with respect to TE. Particularly, considering two echo times such that $TE_1 < TE_2$, we obtain:

$$\omega_{\mathbf{r}} = \frac{\phi(\mathbf{r}, TE_2) - \phi(\mathbf{r}, TE_1)}{TE_2 - TE_1} \,. \tag{2.25}$$

2.4.3 Correcting effects of $\omega_{\mathbf{r}}$

Diverse schemes have been proposed in the literature for correcting the effects of $\omega_{\mathbf{r}}$. In [AP00] and [Lin+12], authors consider the perturbations from $\omega_{\mathbf{r}}$ as applying a convolution, either in image domain with a modified PSF (more details on this in the next Section) or in k-space with the so-called Modulation Transfer Function (MTF), respectively. As the underlying term $\omega_{\mathbf{r}}t$ in Eq. (2.21) is dependent on both spatial and temporal (i.e. k-space) domain, some methods have been developed [Nol+91; MPM97b; SNF03; Fes+05] that involve splitting it into a sum of variables that are each dependent in a single domain.

However, both the above described methods require high resolution ΔB_0 field maps for accurate corrections, which can be estimated only with additional scans. Hence, methods have been explored in the literature that estimate these field maps internally during correction. These internal estimation methods can be optimization driven [MPM97a; SNF04; Pat+20; PWP21] or based on multi-echo imaging sequences [NN00; BN17].

2.5 Analyzing the performances of sampling patterns

It is important to analyze the performances of the designed k-space trajectories and corresponding reconstruction algorithms to understand their potential limitations in terms of image quality and then benchmark their robustness to noise and off-resonance.

2.5.1 Point Spread Function analysis

The first step in analyzing the performance of a subsampled k-space sampling pattern consists in computing its PSF. The PSF is defined as the impulse response of this subsampling system characterized by the k-space sampling pattern. Briefly, it represents the output reconstructed image in response to a hypothetical point source in the image domain. Such PSF analysis is

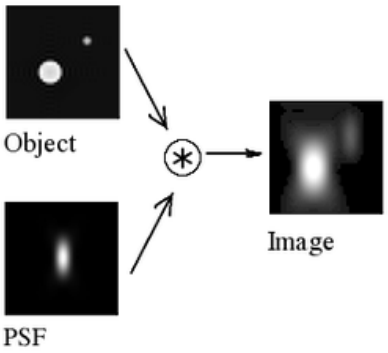


Figure 2.10 – An illustration showing how the ideal reconstructed image can be viewed as the convolution of the PSF of the subsampling pattern with the object image.

important for any imaging system as it affects the spatial mapping of the ideal representation of an object into the observed imaging, which helps to analyze how the imaging method affects image quality, spatial resolution and behaves in presence of noise and off-resonance effects. In an ideal case where MR acquisition process can be modelled as a linear shift invariant system, the PSF of a subsampling pattern can be used to predict the reconstructed image of an object through convolution, see Figure. 2.10[2].

The PSF of a subsampling pattern for an idealistic setting in absence of $\delta B0$ and T_2^* decay can be estimated through adjoint NUFFT operation on the acquired k-space data from the hypothetical point source of the image. The latter being defined as the Dirac delta function in the image domain, its Fourier transform is a constant unity. Then for PSF analysis the acquired k-space data is unity for any location in k-space. However, in the case of non-Cartesian imaging, as described in Section. 2.3.2, the simple adjoint NUFFT of this k-space data would over-represent the densely acquired data points. To avoid this, this constant unity k-space data is precompansated with DCp, through multiplication with the density compensators $\mathbf{w}$ (see [Pau, Sec. 5.5.3]). Mathematically, the PSF of a subsampling pattern Ω can be estimated as:

$$\mathrm{PSF}(\Omega) = F_\Omega^H D \mathbf{1}_\Omega \tag{2.26}$$

with F_Ω^H being the adjoint NUFFT operator, $\mathbf{1}_\Omega$ being the constant unity k-space data and D is the diagonal matrix with $\mathbf{w}$ as diagonal entries (as described in Section. 2.3.2).

Note that while we could use the reconstruction algorithm to reconstruct the PSF, the PSF is a tool used to reflect the quality of the subsampling pattern alone and not the reconstruction algorithms. This is particularly the case when using CS based reconstruction as the reconstruction algorithms enforce some priors of the image being met during reconstruction process. However, the PSF is defined as the impulse response of the subsampling system, hence it involves a point source as input and thus may depart from the priors used on conventional MR images. As a consequence, using a reconstruction algorithm to conduct the PSF would not be fair and would influence the results and estimation of that PSF. This influence is more pronounced in the case of learning based reconstruction algorithms discussed in the next chapter, where the priors learned are more complex and specific to the use case of MRI for a specific organ and possibly a given imaging contrast. The use of PSF is merely a tool to understand the nature of a subsampling pattern, the type of artifacts it may introduce and to assess its robustness to off-resonance and noise through its metrics discussed hereafter.

[2]Source: https://en.wikipedia.org/wiki/Point_spread_function

Simulating effects of off-resonance and T_2^*

A significant use of PSF is that we can do a first level analysis of the k-space sampling trajectory in the presence of off-resonance and T_2^*, by assuming a spatially constant ΔB_0 and T_2^*. For this, we simulate the effects of a spatially constant off-resonance given by $\omega_{\mathbf{r}} = \omega_c$ and T_2^* given by $\alpha_{\mathbf{r}} = \alpha_c$ on the potentially acquired k-space data $\mathbf{1}_\omega$. Assuming the k-space sampling pattern Ω with trajectories $\mathbf{k_i}[n]\ \forall\ i \in \{1, 2, \dots N_c\}, n \in \{1, 2, \dots N_s\}$ of N_c shots and N_s samples per shot, the simulated k-space data $\mathbf{y} = (\mathbf{y}_i[n])_{i=1}^{N_c},\ \forall n \in \{1, 2, \dots, \lfloor \frac{T_{\text{obs}}}{\delta t} \rfloor\}$ is given by:

$$\mathbf{y}_i[n] = \mathrm{e}^{-(\alpha_c + i\omega_c)(n\delta t + t_0)}, \quad \begin{matrix} \forall i \in \{1, 2, \dots N_c\} \\ \forall n \in \{1, 2, \dots, \lfloor \frac{T_{\text{obs}}}{\delta t} \rfloor\} \end{matrix} \tag{2.27}$$

where δt is the dwell time of ADC and t_0 is the time of acquiring the first shot after RF excitation. Then the simulated PSF is given by $\mathrm{PSF}(\Omega, \omega_{\mathbf{r}}, \alpha_{\mathbf{r}}) = F_\Omega^H D\mathbf{y}$.

PSF metrics

Having estimated the PSF, we define some metrics to characterize and quantify the PSF which reflects the performance of the subsampling pattern. We also show these metrics in Figure. 2.11 for clearer understanding.

- **Full Width at Half Maximum (FWHM)**: We define FWHM as the width of the main peak of the PSF at half of its maximal value. This metric is a measure of the peakiness of a given PSF. In order to have a clear sharp reconstructed image, it is important to have a peaky PSF with a small FWHM. In an ideal imaging scenario, the PSF would be a Dirac pulse with FWHM of zero.
- **Peak-to-Sidelobe Level (PSL)**: The PSF of most sampling patterns consist of the main peak and a number of sidelobes. The PSL is defined as the ratio of the main peak to the maximum of the sidelobes. This metric is also a measure of peakiness of the PSF as it measures the relative strength of the main peak with respect to sidelobes. For good imaging, the PSL should be high, and it is unbounded for positive improvements.
- **Peak-to-Noise Level (PNL)**: The subsampling of k-space results in a noise like characteristics in the PSF at locations further from the central peak. Note that this noise is purely from subsampling and is not associated with the acquisition noise in k-space. The PNL is defined as the ratio of the main peak to the noise floor level and inherently measures the level of incoherent noise like artifacts in the reconstructed image. Again, for good imaging, the PNL should be high, and it is unbounded for positive improvements.

Transform point spread function

While the above discussed PSF is a natural tool to measure the performance of a subsampling pattern, it does not account for the sparsity priors enforced in a sparsifying domain through Ψ and corresponding CS based reconstruction schemes as discussed in Section. 2.1.3. For this the notion of PSF was generalized to Transform Point Spread Function (TPSF) in [Lus+07], which measured how a single coefficient in sparse domain is influenced and influences other coefficients through under-sampling in the measurement domain. In the particular case of 2D these coefficients are defined as:

$$\mathrm{TPSF[i;j]} = e_j^* \Psi F_\Omega^H F_\Omega \Psi^H e_i \tag{2.28}$$

where e_i and e_j are unit coefficients in the sparse domain. Characterizing the TPSF and measuring its sidelobes can be helpful in measuring the level of incoherence in the subsampling pattern.

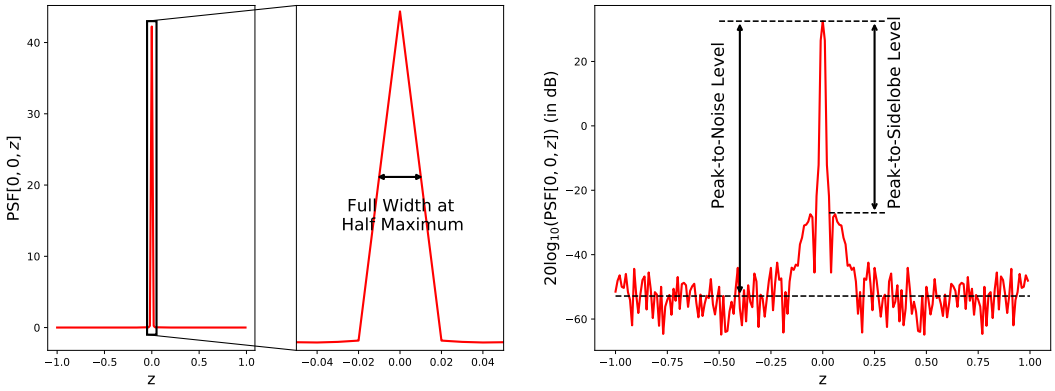


Figure 2.11 – A sample figure showing the PSF of a hypothetical subsampling pattern. The PSF along the z line in mid x and y planes is shown on left with a zoom in to show the FWHM. We represent the same line plot in log scale, scaled to dB on the right and highlight the PSL and PNL in the plot.

2.5.2 Image quality metrics

A large variety of image quality metrics have been proposed in literature to evaluate and quantify the performance of subsampling pattern and corresponding reconstruction algorithms. However, most of these metrics rely on the knowledge of the ground truth image, which is not available in practice always. Furthermore, these ground truth images must be acquired with the same acquisition parameters as the subsampled k-space data in the same orientation without any motion of the patient, which is not always possible. This prevents the use of these metrics in *prospective* evaluations at scanner as ground truth cannot be obtained through fully sampled Cartesian acquisitions, which usually suffer from inter scan motion. In such cases, qualitative analysis of the reconstructed images is the only viable option to evaluate performance.

However, the below described image quality metrics can be efficiently used to evaluate the performance in the case of *retrospective* studies, where the acquired k-space data is simulated by using NUFFT. Such retrospective studies can be really helpful to quickly benchmark a variety of trajectories and reconstruction algorithms, and also grid-search the optimal regularization or target sampling density parameters.

In what follows, we describe the image quality metrics between ground truth image $\mathbf{x}$ and the reconstructed image $\hat{\mathbf{x}}$. Usually to both these images are normalized with their maximum or average values to yield consistent metric values in the same range to judge reconstruction performance over varied imaging dataset.

MSE

The most common metric is the L2 norm between the ground truth image $\mathbf{x}$ and the reconstructed image $\hat{\mathbf{x}}$ given by MSE:

$$\mathrm{MSE}(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{N} \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \,. \tag{2.29}$$

Variations of this metric exists like Root Mean Squared Error (RMSE), which is square root of MSE and Normalized Mean Squared Error (NMSE) which normalizes the loss with respect to ground truth and is given by:

$$\mathrm{NMSE} = \frac{\mathrm{MSE}(\mathbf{x}, \hat{\mathbf{x}})}{\mathrm{MSE}(\mathbf{x}, \mathbf{0})} \tag{2.30}$$

Typically, the lower the value of these metrics, the better the reconstructed image is.

PSNR

Inspired by the use of SNR, another metric used to evaluation reconstructed image performance which is a mix between SNR and MSE is PSNR:

$$\mathrm{PSNR}(x, \hat{x}) = 20 \log_{10} \left(\frac{\max(|\hat{x}|)}{\mathrm{RMSE}(x, \hat{x})} \right) \tag{2.31}$$

Notice that PSNR is a logarithmic scale, and inversely propotional to RMSE, hence the higher the value of PSNR, the better the reconstructed image is. Additionally, PSNR metric is unbounded for any improvement in the reconstructed image quality.

SSIM

Some advanced metrics like SSIM [Wan+04] have been proposed in literature which is based on human visual perception of images and characterized through several indices like luminance l, contrast c and structure s defined as follows:

$$l(\mathbf{x}, \hat{\mathbf{x}}) = \frac{2\mu(\mathbf{x})\mu(\hat{\mathbf{x}}) + c_2}{\mu^2(\mathbf{x})\mu^2(\hat{\mathbf{x}}) + c_2} \tag{2.32}$$

$$c(\mathbf{x}, \hat{\mathbf{x}}) = \frac{2\sigma(\mathbf{x})\sigma(\hat{\mathbf{x}}) + c_2}{\sigma^2(\mathbf{x})\sigma^2(\hat{\mathbf{x}}) + c_2} \tag{2.33}$$

$$s(\mathbf{x}, \hat{\mathbf{x}}) = \frac{\sigma(\mathbf{x}, \hat{\mathbf{x}}) + c_3}{\sigma(\mathbf{x})\sigma(\hat{\mathbf{x}}) + c_3} \tag{2.34}$$

with μ the averaging function, σ the standard deviation and $\sigma^2(x, \hat{x})$ the covariance of x and $\hat{x}$, and c_1, c_2, c_3 constants relative to maximum image values as explained in [WSB03b].

From [Wan+04], the SSIM is defined as:

$$\mathrm{SSIM}(x, \hat{x}) = l(x, \hat{x})^{\alpha} \times c(x, \hat{x})^{\beta} \times s(x, \hat{x})^{\gamma} \tag{2.35}$$

where the parameters α, β, γ are coefficients to weigh the importance of each term and are all set to 1 in [Wan+04].

An extension to SSIM is multi-scale SSIM [WSB03b], which proposes to analyze the image at multiple scales M_S:

$$\mathrm{MSSIM}(\mathbf{x}, \hat{\mathbf{x}}) = l(\mathbf{x}, \hat{\mathbf{x}})^{\alpha_{M_S}} \prod_{j=1}^{M_S} c(\mathbf{x}, \hat{\mathbf{x}})^{\beta_j} \times s(\mathbf{x}, \hat{\mathbf{x}})^{\gamma_j} \tag{2.36}$$

where all the weighing factors α_j, β_j, γ_j are set to 1 for simplicity in [WSB03b]. Throughout our experiments, we used the default values for these weighing factors from their implementation functions in SciKit-Image and TensorFlow.

2.6 Towards the use of deep learning

Throughout this chapter, we have summarized the core concepts of CS in MRI, the need for VDS which is achieved efficiently through non-Cartesian imaging. Particularly, we discussed how to undersample the k-space trajectories and discussed a variety of reconstruction algorithms to obtain image from this k-space data. However, in recent years, there has been a paradigm shift in the field of MRI from the conventional CS to using DL based reconstruction algorithms. This trend is on rise ever since MR imaging datasets like fastMRI [Zbo+18] and the Calgary dataset [Sou+18] have been made publicly available. As these datasets also contain the raw acquired k-space data, they have been instrumental for benchmarking not only DL based MR image reconstruction algorithms but also data-driven learning based frameworks to optimize the k-space sampling patterns.

In the next chapter, we review some fundamentals of DL and review some literature on DL based reconstruction networks and its rising use in k-space sampling pattern design.

* * *
* *
*

Chapter 3

Deep Learning for MRI

BRIEFLY, from earlier chapters, we can break down the problem of speeding up Magnetic Resonance (MR) scan into two major sub-problems:

- **Sampling pattern design:** How to under-sample the k-space and optimally acquire data through constrained k-space sampling trajectories.
- **Reconstruction:** Having under-sampled the k-space, how to reconstruct the corresponding MR image from this under-sampled data.

The latter problem can be viewed as an ill-posed inverse problem. In recent years, there has been a strong propensity to tackle such problems using DL, which helps design parameterized nonlinear models for a specific target task like regression, classification or segmentation. Later, the corresponding parameters of these models can be calibrated or learned from a given dataset. In the context of inverse problems, such carefully crafted and learned nonlinear models can be used as good priors over the data, which can be helpful in extending the Compressed Sensing (CS) based methods described in previous chapter to obtain improved image reconstruction quality.

In this chapter, we focus on the fundamentals of DL, its formalism and the key idea of backpropagation which allows for optimizing large and complex networks to model intricate dynamic systems. Later, we review some literature on learning of Magnetic Resonance Imaging (MRI) reconstruction networks. More recently, with the rise of publicly available raw k-space datasets, DL has also been applied to learn optimal k-space sampling patterns within a given scan time budget at a given target contrast. We therefore summarize these new methods (e.g. PILOT, BJORK) to learn such k-space sampling trajectories.

We emphasize that during this thesis, the discussed reconstruction networks are purely applied, and we do not develop new methods or update current ones in the literature. However, the core contributions and novelty of this thesis involves learning the under-sampling patterns in the form of physically compliant k-space trajectories jointly along with the corresponding reconstruction networks.

3.1 Deep Learning fundamentals

In this section, we briefly review the fundamentals of DL and particularly focus on back-propagation which is a method to obtain the gradients of a loss with respect to network parameters. Later, we discuss some basic architectural blocks present in the literature which can be combined to yield a complex structure that can be eventually used to model a highly nonlinear function. Finally, we discuss how to learn the network parameters for such complex models using the gradients of the loss function.

3.1.1 Formalism and Backpropagation

DL involves approximating a function f, which maps inputs x to outputs y, using composition of simple parameterized and nonlinear functions, leading to a complex model f_θ which is highly parameterized by θ. These parameters θ are *stochastically* optimized or "learned" during the training step, wherein the parameters are iteratively updated to minimize a target loss function $\mathcal{L}$ with respect to a training dataset $\mathcal{D}$ of length N_T which consists of a set of input and output data pairs given by $(x_i, y_i)\ \forall\ i \in \{1, \ldots, N_T\}$. Such a network is built and optimized in order to model and learn an unknown or partially known process. In some situations, it is done to structure a currently known model using a strongly parallelized connected network to speed up its computations on Graphical Processing Unit (GPU). Mathematically, this learning process can be written as the following optimization problem:

$$\widehat{\theta} = \arg\min_{\theta} \sum_{i=1}^{N_T} \mathcal{L}(f_\theta(x_i), y_i) \tag{3.1}$$

For a basic sequential neural network consisting of N_L composition of simpler functions called "layers", the function f_θ can be written as:

$$f_\theta = f_{N_L} \circ f_{N_L - 1} \circ \cdots \circ f_2 \circ f_1 \tag{3.2}$$

In order to solve the optimization problem in Eq. (3.1), we need to compute the gradients of the loss $\mathcal{L}$ with respect to the network parameters θ. The backbone of obtaining such gradients efficiently for composed architectures given by Eq. (3.2) is the backpropagation algorithm. The gradients associated with the loss $\mathcal{L}$ with respect to a set of network parameters θ_n associated with the intermediate layer f_n can be computed using the chain rule as:

$$\frac{\partial \mathcal{L}}{\partial \theta_n} = \frac{\partial \mathcal{L}}{\partial f_n}\frac{\partial f_n}{\partial \theta_n} = \frac{\partial \mathcal{L}}{\partial f_{N_L}}\frac{\partial f_{N_L}}{\partial f_{N_L - 1}} \cdots \frac{\partial f_{n+1}}{\partial f_n}\frac{\partial f_n}{\partial \theta_n} \tag{3.3}$$

Notice that all the partial derivatives in Eq. (3.3) except the last one is independent of θ_n. Further, all the these intermediate partial derivatives can be used for gradient computation of the previous layer $n-1$ as:

$$\frac{\partial \mathcal{L}}{\partial \theta_{n-1}} = \frac{\partial \mathcal{L}}{\partial f_n}\frac{\partial f_n}{\partial f_{n-1}}\frac{\partial f_{n-1}}{\partial \theta_{n-1}} \tag{3.4}$$

This process can be repeated iteratively as we calculate gradients for the last layer f_{N_L} to the first layer f_1 and is called backpropagation as the gradients are propagated backwards from the last layer to the first layer. Such first-order gradient based optimization methods is used commonly in DL, even though in practice some underlying composing layers or functions may not be differentiable.

3.1.2 Optimization and learning

Once the gradients of loss $\mathcal{L}$ are computed for network f_θ with respect to the network parameters θ, one can use simple gradient descent based algorithms to iteratively update the network parameters. However, note that as the loss in Eq. (3.1) is a sum of loss over the entire training dataset $\mathcal{D}$, the gradient of the loss with respect to the network parameters θ is also a sum of gradients over the entire training dataset. This can be computationally expensive as the gradients are computed for each training sample and then summed up. Further, in the case of implementations using GPU, the entire dataset $\mathcal{D}$ has to be loaded into the GPU memory which can be a bottleneck for large datasets. To reduce the memory and computational cost, one can use a Stochastic Gradient Descent (SGD) algorithm which uses a subset of the training dataset $\mathcal{D}$ to compute the gradient. This is done by randomly sampling a subset of the training dataset $\mathcal{D}$ of length N_B called a "batch" and computing the gradients of the loss with respect to the network parameters θ using this batch. The batch size N_B is a hyperparameter that can be tuned to achieve a good trade-off between the computational cost and the quality of the optimization. The network parameter update for the SGD algorithm can be written as:

$$\theta_{t+1} = \theta_t - \eta \sum_{i=1}^{N_B} \frac{\partial \mathcal{L}(f_\theta(x_i), y_i)}{\partial \theta} \tag{3.5}$$

where η is the learning rate of the algorithm which needs to be appropriately tuned to have fast and stable convergence rate. Such a SGD algorithm shows a trade-off between the computing time and the accuracy of the gradients, controlled by the batch size N_B. In practice, the accuracy of the gradients determines the maximum allowable learning rate η to ensure stable convergence, thereby affecting the convergence rate.

The above SGD update is the simplest form of SGD and is called the "vanilla" SGD. There are many other variants of SGD which are used in practice to achieve faster convergence. The most commonly used variants use "momentum" to track the previous descent directions [Sut+13]:

$$v_{t+1} = \beta v_t + \eta \sum_{i=1}^{N_B} \frac{\partial \mathcal{L}(f_\theta(x_i), y_i)}{\partial \theta} \tag{3.6}$$

$$\theta_{t+1} = \theta_t - v_{t+1} \tag{3.7}$$

where β is a hyperparameter that weighs the current gradient compared to earlier accumulated gradients. More advanced methods rely on normalizing the gradients like adaptive gradient descent (AdaGrad) [DHS11] and adaptive moment estimation (Adam) [KB14]. However, Adam is sometimes unstable and hence in order to accurately learn reconstruction networks, Rectified Adam (RAdam) [Liu+19], which is a more stable variant of Adam with faster convergence, has been widely used.

3.1.3 Architectural blocks

In this subsection, we briefly discuss some commonly used architectural blocks in DL which are used to build the network f_θ. These linear and nonlinear basic building blocks are then exploited to build more complex networks through composition (see Eq. (3.2)) and are disseminated in many applications.

Nonlinear layers

An important aspect of DL is the use of "activation" functions, which are used to introduce non-linearity in the network. Typically, such activation functions map the input to the output through a pointwise nonlinear function where the function is applied independently for all the inputs to produce outputs of the same dimension. For a more extended discussion on non-linearities we refer the reader to the review of [Nwa+18].

The common activation functions used in DL are (see Figure. 3.1):

- **ReLU and variants:** One of the most commonly used activation functions is Rectified Linear Unit (ReLU) [NH10] which is defined as the following pointwise function:

$$\mathrm{ReLU}(x) = \max(0, x) \tag{3.8}$$

  This basic non-linearity allows the network to select a subset of the input features and ignore the rest. However, note that the ReLU function is not differentiable at $x = 0$, while this is usually ignored in practice. Further, the gradients of the ReLU function are zero for negative inputs, which can cause the network to have the so called "vanishing gradient" problem.

  To address this issue, variants of the ReLU function have been proposed in the literature. Some commonly used variants include Leaky ReLU (LReLU) [MHN13] function which is defined as:

$$\mathrm{LeakyReLU}(x) = \max(\alpha x, x) \tag{3.9}$$

  where $\alpha \geq 0$ is a hyperparameter which controls the slope of the function for negative inputs. An extension to LReLU function is the Parametric ReLU (PReLU) [He+15] function where the slope α is learned during training and can be a vector of length equal to the number of input features for an improved control. Many other variants of ReLU have been proposed in the state of the art [RZL17; Zhe+15].

- **Sigmoid, Softmax and Tanh:** Some DL neural networks are used for classification between 2 classes, in which case the output of the network must be as discretized as possible and lie bounded between two values signifying 2 classes (like 0 and 1). In such cases, the sigmoid nonlinear function, which is defined as follows, is used:

$$\sigma(x) = \frac{1}{1 + \mathrm{e}^{-x}} \tag{3.10}$$

  The sigmoid function is differentiable and has a smooth gradient which is useful for training. An extension to this for multi-class classification using a vector of inputs $x \in \mathbb{R}^K$, involves the network outputting a vector of probabilities, for which the SoftMax function is used:

$$\mathrm{SoftMax}(x) = \frac{\mathrm{e}^{x}}{\sum_{j=1}^{K} \mathrm{e}^{x_j}} \tag{3.11}$$

  where K is the number of classes. The SoftMax naming comes from the fact that the function indicates the maximum of the input vector smoothly allowing for differentiability.

  Finally, the tanh function is used to map the input to the range $[-1, 1]$ and is defined as:

$$\tanh(x) = \frac{\mathrm{e}^{x} - \mathrm{e}^{-x}}{\mathrm{e}^{x} + \mathrm{e}^{-x}}\,. \tag{3.12}$$

  However, this function is not used as much in recent years as it is very close to the Sigmoid function, which returns values in the range $[0, 1]$.

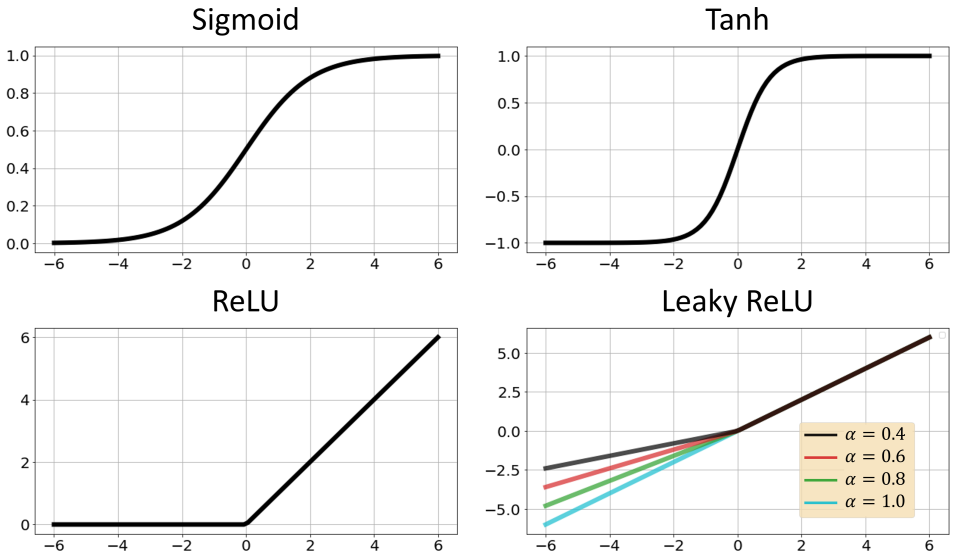


Figure 3.1 – Illustration of common activation functions in deep learning.

Perceptron

The perceptron [Ros58] is the simplest building block of a DL network. It is a linear function which takes a vector of inputs x and outputs a scalar value $f(x)$. Many perceptrons can be chained to form Multi-Layer Perceptron (MLP), which takes a vector of inputs $x \in \mathbb{R}^{p_{inp}}$ and produces a vector of outputs $f(x) \in \mathbb{R}^{p_{out}}$. The perceptron can be written as:

$$f(x) = \sigma(\mathbf{W}x + \mathbf{b}) \tag{3.13}$$

where $\mathbf{W} \in \mathbb{R}^{p_{out} \times p_{inp}}$ is the weight matrix, $\mathbf{b} \in \mathbb{R}^{p_{out}}$ is the bias vector and σ is the nonlinear activation function described in previous subsection. Such a perceptron can also be used to model a linear function $f(x) = \mathbf{W}x + \mathbf{b}$ by setting $\sigma(x) = x$.

Convolutional layers

With the rise of use of DL in computer vision, the need for a more efficient way to process images arose. Conventional MLP networks are not suitable for processing images as they require the input to be a vector, which is not the case for images. While the input to a MLP network can be reshaped to a vector, this is not a good solution as it loses the spatial information of the image. Further, there was a need to learn a function which is equivariant with respect to translation, i.e. the output of the function for a translated input must be the same output translated in the same way:

$$f(\mathcal{T}(x)) = \mathcal{T}(f(x)) \tag{3.14}$$

where $\mathcal{T}$ is a translation operator. This is a desirable property for many computer vision tasks, such as object detection and segmentation.

The Convolutional Neural Network (CNN) [LeC+89; Kri09] was proposed to address this problem, which is now a basic building block of DL networks for imaging related applications. The CNN is a special type of MLP network, where the weights are reshaped to a 2D image and used as a convolutional kernel. The CNN opertaor which takes $x \in \mathbb{R}^{p_{inp} \times h \times w}$ as input and produces $f(x) \in \mathbb{R}^{p_{out} \times h' \times w'}$ as output, where h' and w' are the height and width of the output image, is defined as:

$$f(x) = \sigma(\mathbf{W} * x + \mathbf{b}) \tag{3.15}$$

where $\mathbf{W} \in \mathbb{R}^{p_{out} \times p_{inp} \times k \times k}$ is the weight matrix, $\mathbf{b} \in \mathbb{R}^{p_{out}}$ is the bias vector, σ is the nonlinear activation function and $*$ is the convolution operator.

In practice, the convolution operation is not well defined at the boundaries of the image, where the kernel is not fully contained in the image. To address this point, the most common method involves padding the image with zeros at edges, which is called zero-padding.

Pooling and upsampling layers

Inspired by the wavelet operators, it is important to analyze images at multiple scales to extract features and improve the overall network performance. For this, pooling and upsampling layers are used in conjunction with CNN networks. Pooling is used to downsize an image by reducing the spatial resolution through the following operations:

- **Max pooling**: The maximum value of a window of size $k \times k$ is taken as the output.
- **Average pooling**: The average value of a window of size $k \times k$ is taken as the output.

On the other hand, during upsampling, the image is upscaled to a higher resolution by repeating the pixels in the image by a factor of k.

Normalization layers

Deeper CNN networks are prone to suffer from vanishing and exploding gradients, as the gradients accumulate through backpropagation across such deep layers and can become very small or large. Normalization layers are used to address this problem by normalizing the input to each layer, thereby making the networks more stable. Mathematically, the normalization layer is defined as:

$$f(x) = \gamma \frac{x - \mu}{\sigma(x)^2 + \epsilon} + \beta \tag{3.16}$$

where μ and σ are the mean and standard deviation of the input, γ and β are hyperparameters to control the new mean and variance of the input, ϵ is a small constant to avoid division by zero. Such normalization is applied repeatedly in the network to control the mean and standard deviation of the input to each layer, to explicitly control the energy of the input to the network thereby leading to preconditioning of the corresponding gradients. Such normalization layers are shown [IS15; San+18] to improve the stability, convergence rate and performance of the network. In practice, such normalization is applied along a dimension of the input, and many variants of such normalization layers exist, such as batch normalization [IS15], layer normalization [BKH16] and instance normalization [UVL16].

Residual and skip connections

While it is possible to learn a function f by stacking a large number layers of nonlinear functions, the output of such a large and deep network can lose coherence with respect to the input for out-of-distribution cases. A way to tackle this issue is to have some skip connections where the features of a layer are concatenated with the inputs of the network for next layer, so that some coarse information of the input exists. This way, the backpropagation can be split into a deep and a shallow path, where the shallow path can learn the coarse information of the input and the deep path can learn the fine details of the input.

Another way to tackle this issue is to use residuals, where the network is learned to output the difference between the inputs and targets, instead of the targets directly:

$$f(x) = x - f_\theta(x) \tag{3.17}$$

Here we learn the function $f_\theta(x)$, which is parametrized by θ, to model the difference between the input x and the target $f(x)$. Such residual connections are particularly useful for image reconstruction tasks where the input and target are similar, and the network is expected to learn the difference between the two.

Dropout and regularization

As the number of trainable parameters grows with the depth of the network, the network is prone to overfitting the training data. To address this issue, regularization is carried out where the network penalties are penalized for large weights through addition of a regularization term (usually the L2 norm of the weights) to the loss function.

Another popular method to address overfitting in DL is to use dropout [Sri+14], where a fraction of the neurons are randomly dropped out during training. This way, the network is forced to learn robust features which are not dependent on a particular neuron.

Finally, data augmentation is used to increase the size of the training data by applying random transformations to the training data, such as random cropping, random rotation, random flipping, etc. This way, the network is forced to learn robust features which are not dependent on a particular transformation of the input.

3.1.4 Universal approximation theorem

The goal of DL is to approximate the unknown function f through learning of network parameters θ of the model f_θ. This raises the question if the network of nonlinear and parameterized functions as shown in Eq. (3.2) can approximate any function f. To answer this question, some theoretical results [HSW89; AZ17; Han19] have been obtained in the mathematical literature of DL, which states that neural networks are universal function approximates, i.e. sufficiently large and deep network can approximate any function f arbitrarily well. However, the underlying theory still lacks to fully understand the dynamics of such networks and is still an active area of research in theory of DL.

The above result of universal approximation of neural networks can be generalized to other types of networks like CNN [Mar+19; Yar21] which is translation-equivariant network commonly used for image domain applications. In practice, the performance of such networks increases with the depth of the network [Tel16; Pez+20b], as this drastically improves the flexibility of the network to model complex functions.

3.1.5 Limitations of DL

DL is a powerful tool for learning complex functions, however, it is not a panacea for all problems. The theory of DL is evolving rapidly as it is subject to intensive research. Therefore there are many open problems in the field:

- **Non-Convexity:** The optimization problem in Eq. (3.1) is non-convex and the optimization can get stuck in local minima. Due to this, the initial parameters of the network can have a significant impact on the final performance of the network [TF95; PP04; Sou16]. This raises a need to open source the codes along with the seeds used for any random initialization of network parameters in all scientific communications, so that the community can reproduce the results and compare them with other methods.

- **Generalization and Overfitting:** The objective of DL is to learn the underlying function f from a set of training data pairs (x_i, y_i). However, as discussed earlier, as the performance of network increases with increase in the depth of the network, there is a propensity to over-parameterize the architectures that fit the training data well, but fail to generalize and model the function f well thereby leading to poor performance on out-of-distribution data. This phenomenon is called overfitting and to overcome it the training data is often accompanied by a validation set to monitor the performance of the network on unseen data.

- **Interpretability:** The network parameters θ are often not interpretable, and it is difficult to understand the underlying function f from the network parameters θ. Due to this, the network parameters θ are often not used for clinical decision-making.

- **Computational Complexity:** The computational complexity of training a neural network is often high and the training time increases with the increase in the depth of

the network. Particularly, for the case of reconstruction multi-channel 3D MRI data, the training time can be days to weeks. Further, as the training is usually carried out on GPU, the network depths are limited by amount of GPU memory available. Downsizing the network to fit into memory reduces the expressivity of the network thereby underfitting the data.

- **Proxy loss functions:** The choice of loss function is often subjective and depends on the application and such a choice affects the overall performance of the network. However, a key requirement of this loss $\mathcal{L}$ is that it must be differentiable to be able to use gradient based optimization methods. This may not be always possible as the associated loss for some target applications could be extremely complex and even subjective without any well-defined mathematical formulation. This is particularly true for the case of MRI reconstruction where the performance of networks is subjective and based on organ, contrast and depends strongly on the diagnostic utility of the image. However, such complex criteria cannot be formulated and in practice image quality metrics between the ground truth and the network output are used as a proxy instead which may not perfectly fit the given task at hand.

- **Dataset and biases:** The performance of DL is often dependent on the dataset used for training, and it is crucial to have a large dataset which spans all the important use cases. Additionally, some features in the dataset may be over-represented and bias the solution [Tom+17].

3.2 Learning for MRI reconstruction

In recent years, there has been a paradigm shift in applying DL based methods for solving inverse problems, as such methods can learn stronger and more complex priors from a large dataset of training data, thereby outperforming the traditional variational methods that rely on hand crafted priors (e.g. total variation or wavelet-based regularization). Ever since the availability of large MR datasets like fastMRI [Zbo+18], there has been a surge in the number of DL based methods for MRI reconstruction. The main reason for this is that such DL based methods can be used to model the complex priors from the k-space data, allowing for more accurate reconstruction of the image as compared to sparsifying transforms as used in CS based methods. Further, such priors can be used to remove the artifacts in the reconstructed image caused by k-space under-sampling, thereby allowing for reaching higher acceleration factors without compromising image quality.

A large variety of methods exists in the literature to learn these priors from data and later also using them during the reconstruction process. Broadly, these methods can be categorized into models which learn under constrained setting such that the solutions satisfy the underlying MR physics, and models which are physics-blind and learn the priors from the dataset without any constraints.

In what follows, we discuss these methods to learn a network f_θ parametrized by θ, to reconstruct an image $\widehat{\mathbf{x}}$ from the k-space measurements $\mathbf{y}$. We assume we have a training dataset $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i)\}$ with N_T data points, where $\mathbf{x}_i$ is the ground truth MR image obtained by fully sampling the k-space and $\mathbf{y}_i$ is the undersampled k-space measurements corresponding to the image $\mathbf{x}_i$.

3.2.1 Physics-blind methods

In the initial years of DL based methods for MRI reconstruction, the methods were physics-blind and the DL architectures were versatile with the goal to reconstruct the image purely based on the measurements in the k-space. Such methods ignore the underlying MR physics described in the past two chapters and solely treat the problem as solving a generalized ill-posed inverse problem.

During the training stage, the network is trained to minimize the reconstruction loss $\mathcal{L}$ between the ground truth image $\mathbf{x}_i$ and the reconstructed image $\widehat{\mathbf{x}}_i = f_\theta(\mathbf{y}_i)$ from

undersampled k-space measurements $\mathbf{y}_i$, where $(\mathbf{x}_i, \mathbf{y}_i) \in \mathcal{D}$. Mathematically, this can be written as:

$$\widehat{\theta} = \arg\min_{\theta} \sum_{i=1}^{N_T} \mathcal{L}(\mathbf{x}_i, f_\theta(\mathbf{y}_i)) \tag{3.18}$$

where $\widehat{\theta}$ is the optimal set of parameters of the network which minimizes the reconstruction loss $\mathcal{L}$. Later, the trained network is used to reconstruct the image from the k-space measurements as $\widehat{\mathbf{x}} = f_{\hat{\theta}}(\mathbf{y})$. Note that in this method, the input to the network are the k-space measurements $\mathbf{y}$ presented in the Fourier domain and the output is in the image domain.

Such a method was first applied in AUTOmated transform by Manifold APproximation (AUTOMAP) [Zhu+18] where the authors proposed a DL based method to reconstruct the image directly from the k-space measurements. The authors used MLPs as kernels to translate the k-space measurements to the image domain, followed by CNN to refine the reconstructed image. As the method was completely blind to MR physics and the undersampling patterns involved, it could be applied directly for both Cartesian and non-Cartesian sampling patterns. Further, this method could also be directly extended to other medical imaging modalities such as Computed Tomography (CT) and Positron Emission Tomography (PET). However, the underlying MLP kernels did not scale well for high resolution and 3D MRI data. This was particularly the case as these MLPs learned the mapping from Fourier domain to image domain as pseudo Discrete Fourier Transform (DFT) matrices, which are computationally expensive. Knowing that fast implementations for DFT already exists in literature (i.e. Fast Fourier Transform (FFT)), it was natural to use them in the network architecture itself, thereby learning to minimize the loss $\mathcal{L}$ in a single domain.

Such models can be briefly categorized based on whether they worked in the image domain or the k-space domain:

$$\hat{\theta} = \arg\min_{\theta} \sum_{i=1}^{N_T} \mathcal{L}(\mathbf{x}_i, f_\theta(F_\Omega \mathbf{y})) \tag{3.19}$$

$$\hat{\theta} = \arg\min_{\theta} \sum_{i=1}^{N_T} \mathcal{L}(\mathbf{x}_i, F_\Omega f_\theta(\mathbf{y})) \tag{3.20}$$

where F_Ω is the FFT operator and Ω is the sampling pattern. Notice that Eq. (3.19) learns a network in the image domain while Eq. (3.20) learns a network in the k-space domain.

Both methods are constantly used in the literature where the networks primarily consist of CNN layers [Hyu+18; Lee+18; Han+18; Akç+19; HSY19].

However, similar to wavelets it is important to have CNN layers with different receptive fields to capture the different scales of the image. For this, U-Net [RFB15b] was implemented which consists of CNN kernels operating at different scales. In practice, this is obtained by applying CNN kernels with different receptive fields through downsizing the image using pooling layers. Later, these features are upsampled and concatenated with the features from the corresponding layer at the same scale as skip connections. Such a network is shown in 3.2, and is called U-Net as it looks like a U-shape.

Note that the use of CNN kernels in the image domain is justified due to the desired translation equivariance. However, its application along with non-linearities in k-space [HSY19] can pose issues due to lack of such equivariance and also as the energies in k-space are extremely skewed with most of the signal energy concentrated at the center of k-space. Some models are explored in the literature which extend the linear Generalized Autocalibrating Partially Parallel Acquisitions (GRAPPA) kernel to a nonlinear kernel leading to Robust Artificial neural network for K-space Interpolation (RAKI) [Akç+19].

A major limitation of such physics-blind methods is the lack of data consistency steps, which reduce the recovery guarantees of the reconstructed images. Further, the probability of hallucinations (see Figure. 3.3) is also high in such methods, which is can be detrimental for the use of such methods in medical diagnosis.

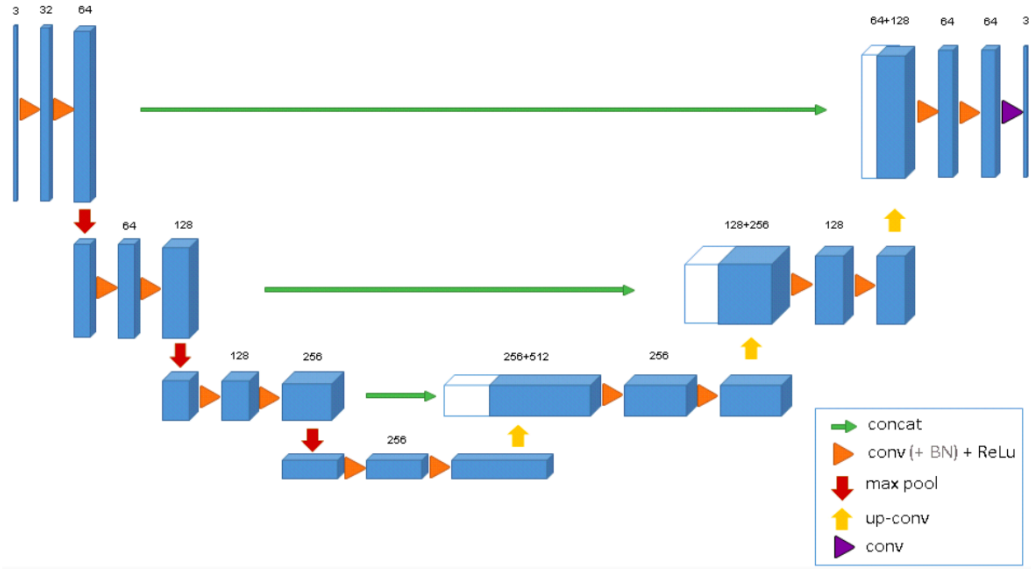


Figure 3.2 – A typical illustration of the U-Net multi-scale network with feature maps in blue, skip-connections in green. From [Çiç+16].

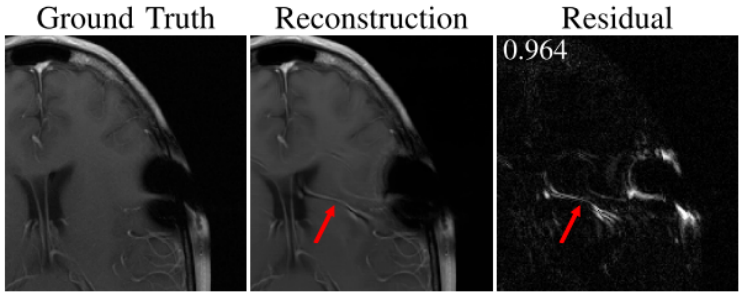


Figure 3.3 – A typical illustration of hallucinations occurring in reconstruction of MR images using DL. From [Muc+21b]. The left image is the original image, the middle image is the reconstructed image with hallucination artifact pointed with a red arrow, the right image corresponding to the residuals, i.e. their absolute difference.

3.2.2 Unrolled Networks

As illustrated earlier, the presence of data fidelity or data consistency is important for stable reconstruction of images with some guarantees. In practice, this is achieved by unrolling the iterative reconstruction algorithms and learning the required priors in the network architecture. This method improves the guarantees of the reconstruction and also reduces the hallucinations in the reconstructed images, as the networks learn to only improve the images with respect to noise and artifacts from under-sampling of k-space.

The most common method to do this involves learning the prior in the form of a regularizer or its proximal operator in the iteration step for CS based reconstruction. With this, the reconstruction iteration step for the single coil case can be written as:

$$\mathbf{x}^{(k+1)} = f_\theta^{(k)}(\mathbf{x}^{(k)} - F_\Omega^H(F_\Omega \mathbf{x}^{(k)} - \mathbf{y})) \tag{3.21}$$

where $f_\theta^{(k)}$ is the proximal network to denoise the image at iteration k. While each of the networks $f_\theta^{(k)}$ at all the iterations must be trained end-to-end [Sch+17; Eo+18] for best performance, in practice for high resolution and 3D multi-coil MRI it is done step-by-step [Ozt+22] to reduce the computational cost. Additionally, to reduce the model's memory footprint, in some methods like MoDL [AMJ19], the network weights are shared

across iterations. The above defined unrolled iteration involves networks to correct the image in a single domain only with data fidelity applied in the k-space domain. A generalized version of this algorithm would also involve the use of networks to correct the image in the Fourier domain [Eo+18], as shown in Figure. 3.4

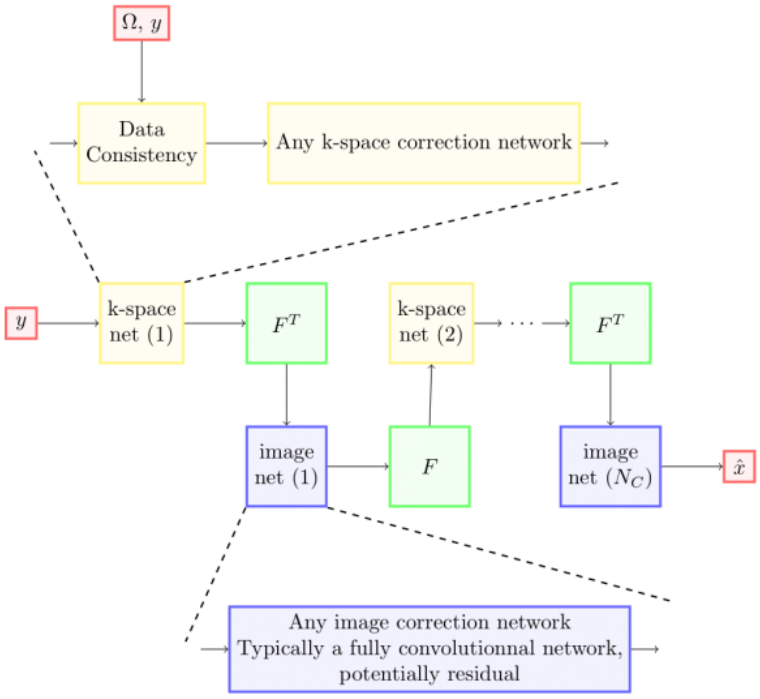


Figure 3.4 – The measurements y acquired following the sampling pattern Ω to obtain the reconstructed image $\hat{\mathbf{x}}$, with F the Fourier transform and N_C the number of iterations. The inter-iteration connections are omitted for clarity. From [RSC21b].

Regularization penalty of $\|x^{(n)} - f_\theta(x^{(n)})\|_2^2$ was added in Model-based Deep Learning (MoDL) by [AMJ18] to reduce the impact of the model and thereby hopefully reducing the amount of hallucinations. More advanced and versatile methods have been proposed in the literature [AÖ18; Ham+19; GOW19; Sri+20] which are inspired through unrolling different iterative algorithms. A popular method is the Primal-Dual (PD) network [AÖ18] which is inspired by the PD algorithm [CP11b]. These algorithms are further accelerated through the use of "memory" by concatenation of past iteration outputs, giving the algorithm an effect similar to momentum in traditional optimization algorithms [BT09; AO17].

These methods were extensively reviewed and benchmarked on the fastMRI [Zbo+18] dataset in [Ram+20] where the PDNet showed to outperform all the other methods. The results of such benchmarking is shown in Figure. 3.6 for knee MR images and Figure. 3.5 for brain MR images.

3.2.3 Reconstructions for non-Cartesian k-space data

Most of the above described physics-based reconstruction methods are designed for Cartesian sampling of k-space data. However, they can be extended to non-Cartesian imaging application by changing FFT operator F_Ω to the Nonuniform Fast Fourier Transform (NUFFT) operator described in previous chapter. Such extension of such a method was carried out in [Sch+19], while it was later noted in [Ram+22c] that such generalization needs density compensation for accelerated convergence and improved reconstructed image quality.

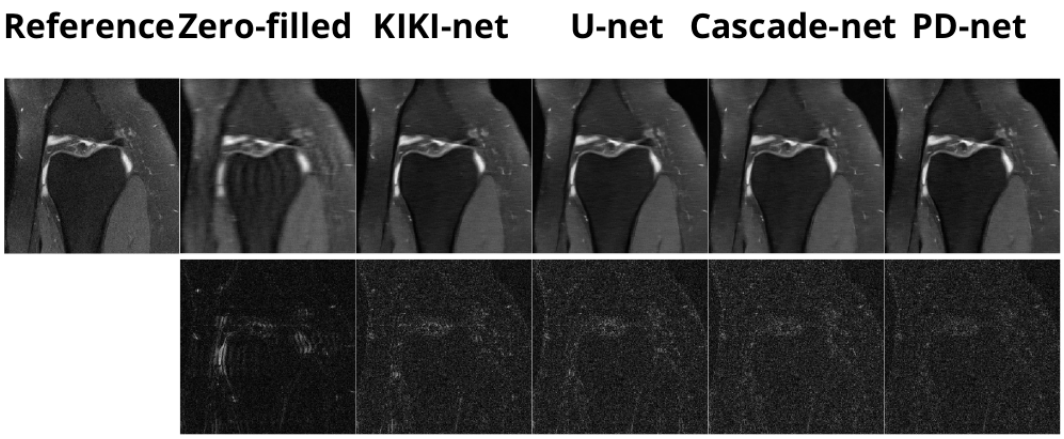


Figure 3.5 – Multiple DL architectures compared one another and to the reference (left) for reconstructing 2D knee images (top row) from Cartesian undersampled k-space data. Absolute errors are shown at the bottom row. The zero-filled case corresponds to the naive reconstruction approach without DL. The best performing architecture is the PDNet. From [Ram+20].

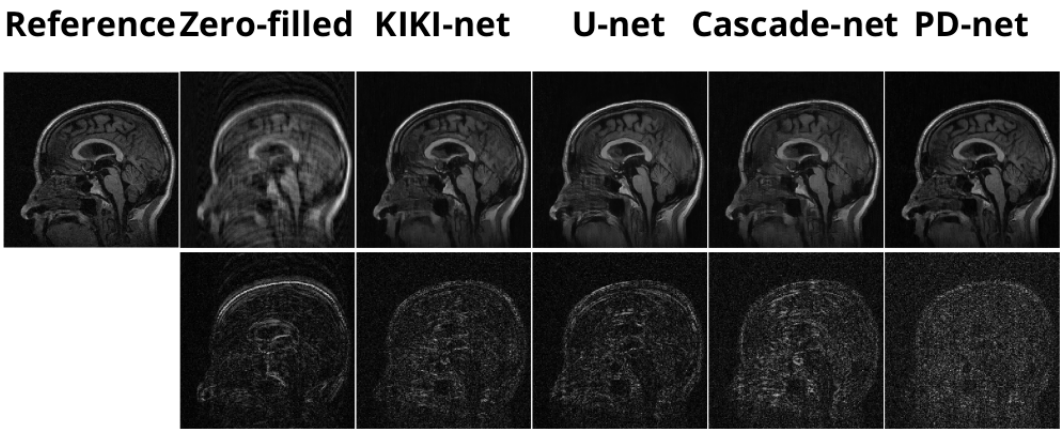


Figure 3.6 – Multiple DL architectures compared one another and to the reference (left) for reconstructing 2D brain images (top row) from Cartesian undersampled k-space data. Absolute errors are shown at the bottom row. The zero-filled case corresponds to the naive reconstruction approach without DL. The best performing architecture is the PDNet. From [Ram+20].

3.3 Learning the sampling patterns

The creation of the fastMRI [Zbo+18] dataset was not only a major step for learning the reconstruction networks, but also for learning the sampling patterns. This is primarily because learning of sampling patterns crucially required the presence of raw k-space data, which did not exist in the community for many datasets previously including Open Access Series of Imaging Studies (OASIS) [LaM+18] and OpenNeuro[1]. While the magnitude images are available, and the k-space is the Fourier domain of these images, note that the images are obtained after a processing pipeline conventionally carried out by the scanner which is not known. Using such methods to infer the k-space data to learn the sampling patterns and also reconstruction networks can produce biased results with inflated performance scores [Shi+22].

With the availability of raw k-space data, for 2D knee and brain MRI in the fastMRI dataset [Zbo+18] and 3D brain MRI in Calgary-Campinas dataset [Sou+18], the community

[1] https://openneuro.org/

has started to explore the learning of sampling patterns. In this section we will review the methods proposed in the literature for learning the sampling patterns. We start by methods which learn Cartesian sampling patterns, followed by methods which learn the gridded Target Sampling Density (TSD) of the k-space. Finally, we discuss some recent strategies to directly learn hardware compliant k-space trajectories. Note that some of these methods did not exist during the course of this thesis, and their overview and comparisons with methods proposed in this thesis are provided in respective chapters (Chapters 6 and 7). We still discuss these methods here for the sake of completeness. Further, the main focus of this thesis lies in non-Cartesian k-space sampling patterns, which are composed of hardware compliant trajectories. While we do present the most important methods to learn gridded under-sampling patterns for Cartesian sampling, we do not go in depth with these methods, as it is beyond the scope of this thesis.

3.3.1 Cartesian sampling

Perhaps the first method to obtain a TSD and corresponding Cartesian sampling patterns was [Kno+11], where the authors used normalized averaged k-space magnitude spectrum of a template MRI dataset as candidate for TSD. This method was purely data-driven and the corresponding optimized sampling patterns did not depend on the type of reconstruction algorithm used. This enforces the sampling points to be densely sampled in regions of high signal energy in k-space, particularly in the center of k-space. While such sampling patterns can be optimal for CS based reconstruction, however improved reconstruction quality can be reached if the sampling was carried out in regions of k-space which cannot be well reconstructed by reconstruction methods.

Taking this into account, in [Göz+18b] a learning-based framework was proposed to learn the subsampling patterns in k-space for a given reconstruction algorithm. The framework is learned through a parameter-free greedy mask selection method which could adapt to a constraint on sampling locations to be in the form of k-space lines, having direct applications in Cartesian MRI. The scaling issues of this algorithm was addressed in [San+19] through a stochastic greedy algorithm, which allowed it to be scaled to 3D parallel MRI and dynamic imaging applications.

As more recent networks try to jointly learn the under-sampling pattern and reconstruction network, in [ZKR21] an alternated learning approach was proposed and applied for parallel MRI. The authors posed a bias-accelerated subset selection algorithm to optimize the sampling pattern. The optimized sampling patterns were shown to exploit the partial conjugate symmetry of the k-space data as discussed in the first chapter.

Later in [She+20], the authors proposed a bi-level learning framework to learn the sparse sampling patterns in k-space and the corresponding regularization weighting. This method was formulated and developed in a supervised setting, using training sets of ground truth images and corresponding k-space measurements. However, the authors showed that only 7 training pairs were sufficient to achieve a high reconstructed image quality.

3.3.2 Learning the sampling policy

Most of the algorithms presented above learned the sampling mask under a given limited scan time budget. However, with a rise in reinforcement learning based methods, there has been increasing interest in the MR community to learn a sampling policy directly. These sampling policies can choose the next sampling location based on the current reconstructed images, and can be used to learn the sampling mask as well. Such methods have wider applications into real time imaging or so-called "active MRI" where the sampling mask is patient specific and dynamic, with scan time promotional to the required reconstructed image quality.

First method in this direction is probably [JUY19], which proposed a self-supervised Monte Carlo Tree Search (MCTS) based approach to learn the "SampleNet" which chooses the next sample position in k-space, along with a corresponding "ReconNet" to reconstruct MR images from the sampled k-space data (Figure. 3.7).

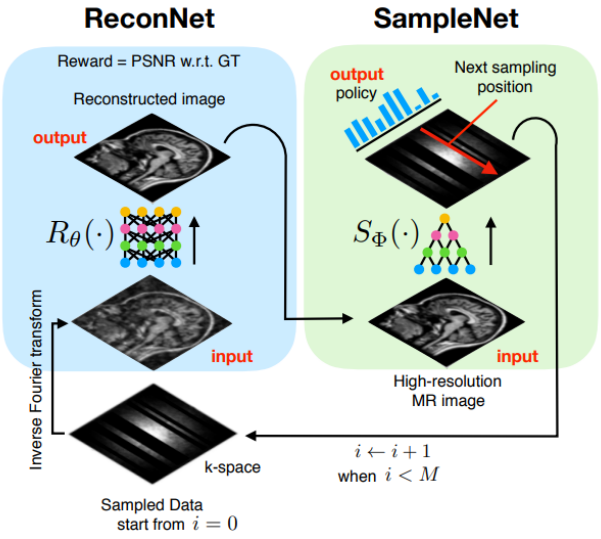


Figure 3.7 – Overall framework of [JUY19] which trains 2 deep neural networks, one to reconstruct the images and the other to estimate a policy to determine the position of the next sample to be collected.

Later in [Pin+20], the task of k-space sampling was modeled as a sequential decision process and used a deep reinforcement learning based approach to learn this sampling policy. In practice, training was performed on large scale fastMRI [Zbo+18] knee dataset using Double Deep Q-Network (DDQN) [HGS15] algorithm.

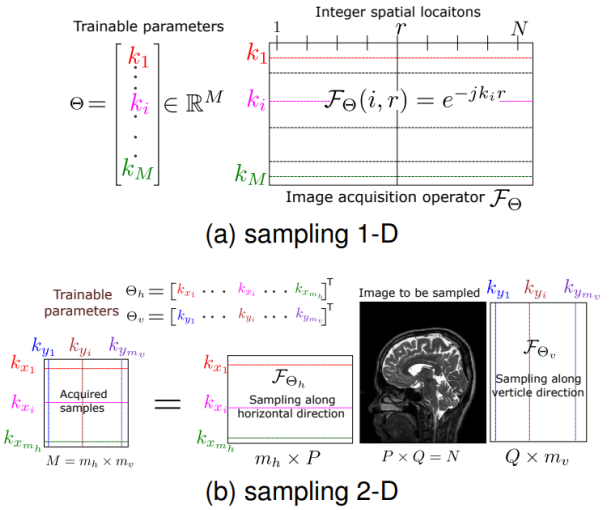


Figure 3.8 – Sampling pattern parametrization in [AJ20] for (a) sampling in 1D parameterized with lines on non-integer locations and (b) sampling in 2D parameterized by horizontal and vertical lines.

Pseudo-Cartesian sampling

All the above methods learn the Cartesian sampling patterns in the form of a gridded mask of binary values indicating the locations of acquired samples. In efforts to extend the algorithm [AJ20] incorporated the NUFFT operator and learned sampling patterns on continuous k-space. However, to tackle the issues of scalability, the sampling patterns were parameterized

in the form of sets of orthogonal lines in the k-space, leading to "pseudo-Cartesian" sampling patterns (Figure. 3.8).

3.3.3 Learning the sampling density

Directly learning non-Cartesian k-space sampling patterns is a challenging and needs more memory, as the samples no longer lie on any Cartesian grid thereby requiring to optimize for the coordinates of the sampling locations. Further, this would need an optimization algorithm where the FFT operators are replaced with NUFFT operators, which are computationally expensive. As an intermediary step, many methods learned the TSD density in k-space, which can be used to draw both Cartesian and non-Cartesian sampling patterns.

To this end, Learning-based Optimization of the Under-sampling PattErn (LOUPE) [Bah+20a] was introduced as a means to learn a probabilistic mask, which was a good candidate for TSD in k-space. The sampling masks were chosen by thresholding the probabilistic mask and the network was trained jointly with the U-net [RFB15b], which was used for reconstructing images from the sampled k-space data. Interestingly, the optimized TSD varied with changes in image anatomy (Figure. 3.9).

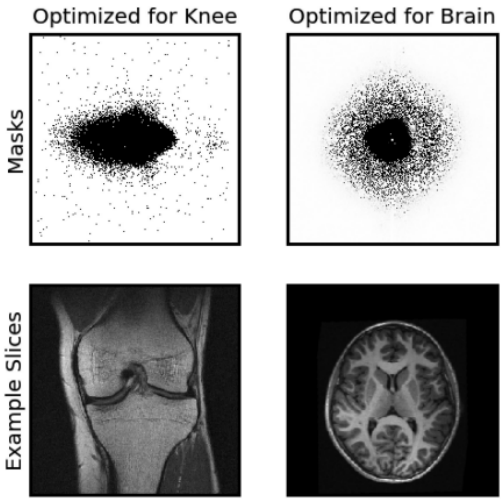


Figure 3.9 – LOUPE-optimized under-sampling masks for under-sampling factor of 8 compared side by side for knee and brain anatomies.

Later, [GGW22] proposed a Bayesian optimization based approach to learn the TSD of k-space. Further, in this method, the authors also optimized the hardware-compliant sampling trajectories through the "sampler", which is an iterative algorithm to minimize the discrepancy [Cha+17], as done in Spreading Projection Algorithm for Rapid K-space sampLING (SPARKLING) [Laz+19].

3.3.4 Non-Cartesian trajectory optimization

With an exponential increase in compute power through GPU and the availability of implementations of the NUFFT operator in frameworks which allowed differentiable programming (like TensorFlow [Mon22; RC23] and PyTorch [Muc+20]), the focus shifted to learning the non-Cartesian sampling trajectories directly. The first work in this direction was Physics-informed learned optimal trajectories (PILOT) [Wei+20] (Figure. 3.10(a)) which produced 2D learned non-Cartesian sampling trajectories and then later extended to 3D [Ved+20] (Figure. 3.10(b)). This method learned hardware compliant k-space sampling trajectories along with U-net [RFB15b] reconstruction network, which was applied on the NUFFT adjoint of the k-space data. Through implementations in PyTorch using TorchKbNufft [Muc+20], this method could backpropagate through the NUFFT operator. However, auto-differentiation

schemes from PyTorch was used to obtain the gradients with respect to k-space sampling locations. These gradients were later shown to be inaccurate in [WF23], which resulted in suboptimality. Later, to overcome this issue, B-spline parameterized Joint Optimization of Reconstruction and K-space trajectories (BJORK) [Wan+21] (Figure. 3.10(c)) was introduced to learn the 2D k-space sampling trajectories using more accurate gradients as described in [WF23].

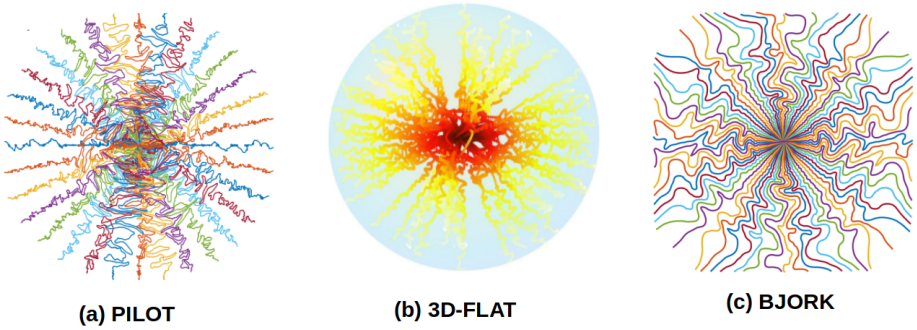


Figure 3.10 – Different methods to learn hardware-compliant k-space sampling trajectories. (a) PILOT [Wei+20], (b) 3D-FLAT [Ved+20] and (c) BJORK [Wan+21].

In order to ensure hardware compliance, all the current state-of-the-art methods enforce penalties in the loss function to penalize the trajectories gradients and slew rates and maintain them below the hardware limits. As the sampling locations are being learned, this leads to many trainable parameters in the overall network, which could lead to suboptimality and convergence to a local minima [Gou+21]. This is tackled in BJORK [Wan+21] through B-spline parametrization of the sampling trajectories, and then multi-scale optimization of the B-spline coefficients similarly to what was originally implemented in SPARKLING [Laz+19].

* * *
* *
*

Part II

Methodological Developments

Chapter 4

Optimizing full 3D SPARKLING trajectories

This chapter covers content that has been published in a peer-reviewed journal:

- **Chaithya, G R**, P. Weiss, A. Massire, A. Vignaud and P. Ciuciu. "Optimizing full 3D SPARKLING trajectories for high-resolution Magnetic Resonance imaging". In: *IEEE Transactions on Medical Imaging* (Mar. 2022)

THE Spreading Projection Algorithm for Rapid K-space sampLING (SPARKLING) al-

gorithm is an optimization-driven method that has been recently introduced for accelerated 2D MRI using compressed sensing. It has then been extended to address 3D imaging using either stacks of 2D sampling patterns or a local 3D strategy that optimizes a single sampling trajectory at a time. 2D SPARKLING actually performs Variable Density Sampling (VDS) along a prescribed target density while maximizing sampling efficiency and meeting the gradient-based hardware constraints. However, 3D SPARKLING has remained limited in terms of acceleration factors along the third dimension if one wants to preserve a peaky Point Spread Function (PSF) and thus good image quality. In this paper, in order to achieve higher acceleration factors in 3D imaging while preserving image quality, we propose a new efficient algorithm that performs optimization on full 3D SPARKLING . The proposed implementation based on fast multipole methods (FMM) allows us to design sampling patterns with up to 10^7 k-space samples, thus opening the door to 3D VDS. We compare multi-CPU and GPU implementations and demonstrate that the latter is optimal for 3D imaging in the high-resolution acquisition regime (600µm isotropic). Finally, we show that this novel optimization for full 3D SPARKLING outperforms stacking strategies or 3D twisted projection imaging through retrospective and prospective studies on NIST phantom and in vivo brain scans at 3 Tesla taking the particular case of T_2*-w imaging. Overall the proposed method allows for 2.5-3.75x shorter scan times compared to GRAPPA-4 parallel imaging acquisition at 3 Tesla without compromising image quality.

4.1 Introduction

The quest for efficient sampling strategies has been a major challenge in MRI since its invention. The theory of Compressed Sensing (CS) [Lus+07] boosted this quest by providing significant theoretical insights. It was proved and observed empirically that for under-sampled acquisitions and approximately sparse signals in an orthogonal basis, an efficient implementation relies on trajectories with a variable density in k-space: the lower frequencies (center of k-space) have to be sampled more densely than the higher at the borders of k-space [Puy+11; Cha+14; Adc+17; Boy+19]. Non-Cartesian k-space trajectories (e.g. spiral, radial, rosette, etc.) [Ahn+86a; Mey+92; Jac+92; Nol97; Law+09; Lus+05a] have been proposed for accelerated and robust-to-motion 2D imaging, prior to the existence of theoretical foundations. While being compliant with scanner hardware constraints on the gradients, these trajectories do not sample the k-space according to a well controlled target sampling density. For instance, in spiral imaging, fulfilling these constraints transforms an initially prescribed density into another one [Cha15, p. 97]. Recently, the SPARKLING algorithm [Boy+16; Cha+17; Laz+19] has been shown to automatically generate optimized non-Cartesian sampling patterns compatible with MR hardware constraints on maximum gradient amplitude and slew rate. SPARKLING optimally samples the k-space (see [Puy+11; Cha+14]) with a controlled distribution of samples (e.g., variable density) and a locally uniform k-space coverage.

However, for the sake of signal-to-noise ratio (SNR), 3 dimensional (3D) imaging is preferred to reach isotropic high-resolution imaging (e.g. 600µm isotropic). In this regard, multiple approaches have been utilized to efficiently down-sample 3D k-space. Some of these involve a combination of a readout in the z-direction with a 2D under-sampled mask based on Poisson disk sampling [Vas+10]. Additional attempts on full 3D readouts were proposed like 3D radial trajectory [Lar+08], 3D cones [Ira+95], twisted projections [Boa+97] and hybrid radial-cones [Joh17]. However, these trajectories were primarily based on parameterizing a k-space sampling curve, and the final sampling pattern was not optimized with respect to the reconstruction quality. Some recent studies explored how to optimize the sampling pattern [Dal+04; Mir+04; Kum+08], but did not include a clear sampling criterion in order to maximize the image reconstruction quality.

Other methodologies in the literature involved stacking a 2D under-sampled trajectory like stack of stars [Son+04; Loo8], stack of spirals [Ira+95; The+99] and stack of 2D SPARKLING [Laz+20a]. Uniform (i.e. cylindrical) stacking is usually implemented even though a spherical strategy, with a number of shots varying as a function of the latitude plan, was shown to be beneficial on image quality for SPARKLING trajectories [Laz+20a].

Further in [Laz+20a], a local 3D SPARKLING approach was proposed by designing a single trajectory within a cone obtained from a parcellation of the 3D spherical k-space. Then all the cones covering a given elevation plane were filled up using the replication of the resulting trajectory. However, this method did not ensure a locally uniform sampling pattern at the boundaries of cones as the problem was solved locally.

The recent rise of machine and deep learning has impacted the literature on MRI sampling [HK19; See+10; Bal+16; Göz+18a; Bah+20b; She+20; Wei+20; Ved+20]. These approaches rely on supervised learning techniques, which means that they need a ground truth corresponding to fully sampled data (like the fastMRI dataset [Zbo+18]), to learn an optimal under-sampling pattern, whether it is Cartesian or not. In [HK19; See+10], the authors explore the use of experimental design to choose the best subset of prescribed trajectories. Although there are substantial differences between these two methodologies (deterministic vs Bayesian, offline vs online design, etc.) they share a similar theoretical background with ours in that sparsity is the key underlying hypothesis. In particular, [See+10; HK19] use the Cramér-Rao bound for sparse signals [BE10] as a tailored optimality criteria. However, such methods are computationally demanding as they try to solve a nonconvex integer programming problem. Hence, in a given time budget this limits the exploration and the potential number of prescribed trajectories.

In [Bal+16; Göz+18a], the authors proposed to step away from the theoretical consideration in CS and adopted a purely data-driven approach. The authors proposed to find an optimal subset of Cartesian sampling lines by using a greedy algorithm aimed at maximizing the SNR. This algorithm can automatically adapt to different reconstruction algorithms and optimality criteria, but its use is limited to Cartesian imaging. More recent approaches made some advance on learning gridded sampling patterns [Bah+20b; She+20]. Additionally, to the best of our knowledge, the only works that have learned a non-Cartesian trajectory under hardware constraints are PILOT [Wei+20] BJORK [Wan+21] and 3D FLAT [Ved+20] for 2D and 3D imaging, respectively. These works seem very promising despite significant theoretical and numerical challenges with a combinatorial number of local minimizers [Gou+21]. In contrast, our work is based on clear theoretical considerations with provable convergence [CB18] in short computing times [MSS21]. Of interest, let us notice that the sampling patterns generated by these methods resemble the SPARKLING ones very much [Wei+20; Wan+21], suggesting that the main ideas behind are now reaching a mature and reliable state. Nonetheless, it is worth noting that none of these approaches has been prospectively validated on real 3D acquisitions. For all these reasons, these works won't be discussed any further in this paper.

In this paper, for the first time, we solve the SPARKLING optimization fully in 3 dimensions. First, in Sec 5.2, we remind the optimization problem to be solved for generating SPARKLING trajectories. Then we focus on major computational bottlenecks that prevented us from scaling the original solution to 3D and provide detail on our main contributions. One key ingredient in SPARKLING is the setting of the right target sampling density. The latter may vary as a function of the resolution, the imaging contrast, the acceleration factor and the object to be scanned. For that purpose, we parameterize radially decaying densities by two parameters (cut-off, decay) and find the optimal density through a grid search over pairs of parameters. This study can be conducted through a retrospective analysis on the target imaging contrast, the organ and given a coil geometry of interest and then the sought optimal density can be used further in prospective acquisitions. For demonstrating the performance of our new trajectories, we use prospective T_2*-w imaging of in vivo human brain. In Sec. 5.4, we present the experimental data sets on which the numerical studies are performed later on for validation and comparison purposes. In this regard, we carry out retrospective and prospective analysis on NIST phantom data collected at 3 Tesla (3T). Then we perform prospective in vivo brain imaging acquisitions on a healthy adult volunteer still at 3T and compare the proposed full 3D SPARKLING with the existing spherical stack of 2D SPARKLING . We do not include any comparison with 3D radial sampling scheme or stack of spirals as this was already done in [Laz+20a]. However, we do compare our trajectories with improved 3D non-Cartesian trajectories, namely twisted projection imaging (TPI) [Boa+97]. TPI trajectories have better k-space coverage as compared to full 3D radial sampling scheme

as these trajectories shift to pappus spirals after a fraction of readout.

4.2 Theory

In this section we briefly introduce the SPARKLING algorithm as described in [Laz+19]. We detail the particular steps involved in the optimization process. We point to some computational bottlenecks in each of these steps. Later, we describe the methods used to overcome these computational challenges, thereby allowing us to scale the problem to 3 dimensions. Most of the theoretical aspects are directly based on earlier works in [Cha+16; Boy+16; Cha+17], which can be consulted for the problem description and derivations of (5.3).

4.2.1 3D K-space sampling

A 3D k-space sampling pattern $\mathbf{K}$ is usually composed of several shots or curves, say N_c, $\mathbf{K} = (\mathbf{k}_i)_{i=1}^{N_c}$, where each 3D shot $\mathbf{k}_i(t) = (k_{i,x}(t), k_{i,y}(t), k_{i,z}(t))$, is controlled by magnetic field gradients $\mathbf{G}_i(t) = (G_{i,x}(t), G_{i,y}(t), G_{i,z}(t))$ as follows:

$$\mathbf{k}_i(t) = \frac{\gamma}{2\pi} \int_0^t \mathbf{G}_i(\tau) d\tau \,, \tag{4.1}$$

with γ the gyro-magnetic ratio ($\gamma = 42.57$MHz/T for proton imaging). In practice, throughout the readout duration T_{obs}, we sample each shot $\mathbf{k}_i(t)$ by a time period Δt, the gradient raster time as the scanner gradient hardware can play gradients at this pace. In the rest of the section, we refer to location of the k-space samples $\mathbf{K}$ as the samples on gradient raster points. Then the number of gradient time steps is given by $N_s = \left\lfloor \frac{T_{\text{obs}}}{\Delta t} \right\rfloor$ and the full 3D sampling pattern $\mathbf{K}$ finally consists of $p = N_c \times N_s$ points. Additionally, we limit ourselves to a long readout ($T_{\text{obs}} \simeq 20$ms) for T_2^*-weighted imaging, as this allows the trajectory to be longer and maximally explore the k-space.

The k-space domain for a 3D MR volume of size $N_x \times N_y \times N_z$ over a field of view $\text{FOV}_x \times \text{FOV}_y \times \text{FOV}_z$, is defined within $[-K_{\max}^x, K_{\max}^x] \times [-K_{\max}^y, K_{\max}^y] \times [-K_{\max}^z, K_{\max}^z]$, with $K_{\max}^\ell = \frac{N_\ell}{2FOV_\ell}$ and $\ell = x, y, z$. For the sake of simplicity, in what follows we assume the same spatial resolution along the three dimensions so $K_{\max}^x = K_{\max}^y = K_{\max}^z = K_{\max}$ even though we meet different matrix and FOV dimensions ($N^z \neq (N^x = N^y)$ and $\text{FOV}^z \neq (\text{FOV}^x = \text{FOV}^y)$). Hereafter, the 3D k-space domain will be normalized to $\Omega = [-1, 1]^3$.

Hardware constraints on the maximum gradient amplitude ($G_{\max}$) and slew rate ($S_{\max}$) induce limitations in trajectory speed and acceleration. These limits can be expressed as box constraints on the amplitude of the discrete derivatives of the k-space trajectory $(\mathbf{k}_i[n])_{n=0}^{N_s-1}$. These hardware constraints can be applied on a per dimension basis, giving rotation variant (RV) constraints, whose resulting trajectories cannot be run on the scanner if the FOV is rotated. Due to this limitation, in this work, we focus on rotation invariant speed and acceleration constraints which can be expressed as follows: [1]

$$\mathcal{Q}_{\alpha,\beta}^{N_c} = \left\{ \begin{array}{c} \forall i = 1, \dots, N_c, \quad \mathbf{k}_i \in \mathbb{R}^{3 \times N_s}, \\ \mathbf{A}\mathbf{k}_i = \mathbf{v}, \\ \|\mathbf{k}_i\|_\infty \leq 1, \ \|\dot{\mathbf{k}}_i\|_{2,\infty} \leq \alpha, \ \|\ddot{\mathbf{k}}_i\|_{2,\infty} \leq \beta, \end{array} \right\} \tag{4.2}$$

where

$$\begin{aligned} \dot{\mathbf{k}}_i[n] &= \frac{\mathbf{k}_i[n] - \mathbf{k}_i[n-1]}{\Delta t} \\ \ddot{\mathbf{k}}_i[n] &= \frac{\mathbf{k}_i[n+1] - 2\mathbf{k}_i[n] + \mathbf{k}_i[n-1]}{\Delta t^2} \\ \|\mathbf{c}\|_{2,\infty} &= \sup_{0 \leq n \leq N_s - 1} \left(|c_x[n]|^2 + |c_y[n]|^2 + |c_z[n]|^2 \right)^{1/2}, \end{aligned}$$

[1] In [Cha+16], we have also dealt with the case of RV constraints where the ℓ_∞-norm replaces the mixed $\ell_{2,\infty}$-norm used here.

for all $\mathbf{c} \in \Omega^{N_s}$ and (α, β) are obtained by normalizing hardware and Nyquist constraints to the sampling domain Ω (from [Laz+19]):

$$\alpha = \frac{1}{K_{\max}} \min\left(\frac{\gamma G_{\max}}{2\pi}, \frac{1}{FOV \cdot \delta t}\right) \tag{4.3a}$$

$$\beta = \frac{\gamma S_{\max}}{2\pi K_{\max}} \tag{4.3b}$$

The purpose of $\mathbf{A}$ and $\mathbf{v}$ are to model linear constraints on the trajectory, like the Echo Time (TE)point constraint, which ensures that each trajectory passes through center of k-space at TE. More sophisticated linear constraints (e.g. gradient moment nulling) can be modeled too, see details in [Cha+16]. The normalized constraint $\alpha\delta t \leq \frac{1}{FOV \times K_{\max}}$ ensures that the distance between k-space locations associated with two consecutive measurements, sampled by the analog-to-digital converter (ADC) at the dwell time period δt (see Subsection 4.4.1 for the relationship between Δt and δt), is lower than the Nyquist rate, which is essential to discard some undesired filtering effects [Laz+20b].

4.2.2 3D SPARKLING formulation

Let $\rho : \Omega \to \mathbb{R}$ denote a target sampling density, with $\rho(x) \geq 0$ for all x and $\int \rho(x)\, dx = 1$. Following previous works [Grä+12; Sch+10; Ehl+19; Laz+19], we obtain $\mathbf{K} \in \Omega^p$ by solving:

$$\widehat{\mathbf{K}} = \arg \min_{\mathbf{K} \in \mathcal{Q}_{\alpha,\beta}^{N_c}} F_p(\mathbf{K}) = \left[F_p^{\mathrm{a}}(\mathbf{K}) - F_p^{\mathrm{r}}(\mathbf{K})\right] \tag{4.4}$$

with $\mathcal{Q}_{\alpha,\beta}^{N_c}$ being the constraint set for the N_c shots. Here we remind that p refers to the total number of k-space samples (or particles), so $p = N_c \times N_s$.

The term $F_p^{\mathrm{a}}(\mathbf{K})$ corresponds to an attraction term which ensures that the final distribution of the k-space sampling points follows the target density ρ and $F_p^{\mathrm{r}}(\mathbf{K})$ is the repulsion term to ensure that the sampling is locally uniform and that we don't have any local clusters. These terms are defined as:

$$F_p^{\mathrm{a}}(\mathbf{K}) = \frac{1}{p} \sum_{i=1}^{p} \int_{\Omega} H(x - \mathbf{K}[i]) \rho(x)\, dx \tag{4.5a}$$

$$F_p^{\mathrm{r}}(\mathbf{K}) = \frac{1}{2p^2} \sum_{1 \leq i,j \leq p} H(\mathbf{K}[i] - \mathbf{K}[j]) \tag{4.5b}$$

where $\mathbf{K}[i] \in \Omega$ describe the locations of k-space samples in a shot-based lexicographical order $[\mathbf{k}_1, \ldots \mathbf{k}_{N_c}]$. The function H is a well chosen kernel, typically $H(x) = \|x\|_2$. Note that alternative choices such as $H(x) = \log(x)$ have been also investigated in [Teu+11]. The minimization problem (5.3) can be attacked by various nonlinear programming procedures. In this work, we propose to use a projected gradient descent as described below:

$$\mathbf{K}^{(t+1)} = \Pi_{\mathcal{Q}_{\alpha,\beta}^{N_c}} \left(\mathbf{K}^{(t)} - \eta^{(t)} \nabla F_p(\mathbf{K}^{(t)})\right) \tag{4.6}$$

The computational bottlenecks in (5.3) involve the calculation of $\nabla F_p(\mathbf{K}) = \nabla F_p^{\mathrm{a}}(\mathbf{K}) - \nabla F_p^{\mathrm{r}}(\mathbf{K})$, and the projection of each shot onto the constraint set $\mathcal{Q}_{\alpha,\beta}^{N_c}$.

4.2.3 Gradient Descent Step

In what follows, we provide details about the calculation of F_p and ∇F_p.

Evaluating F_p^{a} and its gradient

To calculate the attraction term and its gradient, we can re-write (4.5a) as:

$$F_p^{\mathrm{a}}(\mathbf{K}) = \frac{1}{p} \sum_{i=1}^{p} (H \star \rho)(\mathbf{K}[i]) \tag{4.7}$$

where $\star$ denotes the convolution-product in the continuous setting. The main difficulty is thus to quickly evaluate $(H \star \rho)(x)$ (optional, if we want to compute the cost function) and its derivatives. To this end, we discretize the target sampling distribution ρ as follows:

$$\boldsymbol{\rho}[i,j,k] = \rho(i/N, j/N, k/N) \tag{4.8}$$

where $i,j,k \in [-N,N]$, and $N \in \mathbb{N}$ describes the number of discretization points. We typically take N twice as large as $\max(N_x, N_y, N_z)$ to define the density at a better resolution than the image size. Similarly we compute a discrete version of the filter H as:

$$\mathbf{H}[i,j,k] = H(i/N, j/N, k/N) \tag{4.9}$$

Letting $*$ denote the discrete convolution-product, we use the following approximation

$$(H \star \rho)(\mathbf{K}[i]) \simeq \mathcal{I}(\mathbf{H} * \boldsymbol{\rho})(\mathbf{K}[i]), \tag{4.10}$$

where $\mathcal{I} : R^{(2N+1)^3} \to C^0(\Omega)$ denotes a tri-linear interpolant function. Hence, the computation of $F_p^{\mathrm{a}}(\mathbf{K})$ requires to precompute $\mathbf{H} * \boldsymbol{\rho}$ on a discrete grid with fast Fourier transforms once for all. The computation of the sum in (4.7) then has a complexity $O(p)$, which is linear in the number of particles.

Similarly, the computation of $\nabla F_p^{\mathrm{a}}(\mathbf{K})$ involves the calculation of the partial derivatives $\partial_{i,l} F_p^{\mathrm{a}}(\mathbf{K})$ where $1 \leq i \leq p$ is the index of a particle and $1 \leq \ell \leq 3$ the index of a dimension. According to (4.7), the partial derivative is:

$$\partial_{i,\ell} F_p^{\mathrm{a}}(\mathbf{K}) = \frac{1}{p}(\partial_\ell H \star \rho)(\mathbf{K}[i]) \tag{4.11}$$

Thus, letting $\nabla\mathbf{H} \in \mathbb{R}^{(2N+1)^3 \times 3}$ denote a discretization of ∇H, we can precompute the discrete vector field $\nabla\mathbf{H} * \boldsymbol{\rho}$ on a Cartesian grid using fast Fourier transforms and then use a tri-linear interpolant to evaluate it off the grid.

Evaluating F^{r} and its gradient

The problem addressed here is to compute F_p^{r} and $\nabla F_p^{\mathrm{r}}(\mathbf{K})$. For purposes of simplification, we introduce $r_{ij} = \|\mathbf{K}[i] - \mathbf{K}[j]\|_2$ and consider H to be a radial function depending on r_{ij} only. Letting K_ℓ denote the spatial components of $\mathbf{K} = [K_1, K_2, K_3]$, we get:

$$F_p^{\mathrm{r}}(\mathbf{K}) = \sum_{1 \leq i,j \leq p} H(r_{ij}) \tag{4.12a}$$

$$\partial_{i,\ell} F_p^{\mathrm{r}}(\mathbf{K}) = \frac{1}{p^2} \sum_{j \neq i} \left(\frac{K_\ell[i] - K_\ell[j]}{r_{ij}} \right) \partial_\ell H(r_{ij}) \tag{4.12b}$$

The evaluation of all the components of the gradient require $O(p^2)$ computations, where p can reach 10^8 for high resolution imaging. An efficient implementation is therefore critical. In this work, we explored two possibilities.

Brute force calculation using PyKeops The computation of (4.12a) and (4.12b) can be highly parallelized, which is amenable to efficient GPU implementations. Carrying out such computations on array centric frameworks like PyTorch and Tensorflow would require the use of huge $p \times p$-dimensional matrices. This would result in a large memory footprint, much larger than what is typically available on current modern GPUs. For the sake of efficient memory usage, we used `PyKeops`, a library that permits low cost calculations of large kernel operations [Cha+20]. `PyKeops` carries out the naive and direct computations using online map reduce schemes from CUDA routines for summations. Due to this, the whole matrices are not stored in the GPU memory, but rather just the final results.

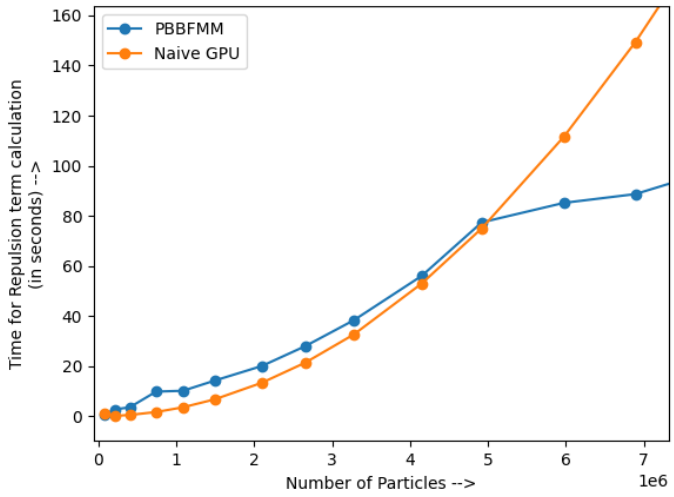


Figure 4.1 – Computation times for the repulsion term F^{r} as a function of the number of particles p.

Fast Multipole Methods Sums of the form (4.12a) and (4.12b) appear in many n-body problems and can be computed efficiently using Fast Multipole Methods (FMM) [Fon+09b]. Given a set of positions $\mathbf{K}[i]$, a kernel $\Psi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ and a set of weights $\mathbf{w} \in \mathbb{R}^p$, the FMM method allows for the efficient computation of vector $\mathbf{b}$ of the form

$$\mathbf{b}[i] = \sum_{j=1}^{p} \Psi(\mathbf{K}[i], \mathbf{K}[j])\mathbf{w}[j] \,.$$

The FMM utilizes a multipole expansion of the kernel Ψ, which allows for a hierarchical grouping of closely spaced k-space points and treat them as a single source. This results in a massive acceleration of the above computation with a complexity $O(p \log p/\epsilon)$, where ϵ is a user-prescribed precision. For our implementations, we used the Parallel Black box FMM [Wan+19; Fon+09a] in 3D (`PBBFMM3D`[2]), which can be run with any arbitrary kernel Ψ.

To evaluate the cost function F_p^{r}, we only need to set

$$\Psi(\mathbf{K}[i], \mathbf{K}[j]) = H(r_{ij}) \text{ and } \mathbf{w}[j] = 1, \forall j.$$

To evaluate the gradient $(\partial_{i,\ell} F_p^{\mathrm{r}}(\mathbf{K}))_i$, we set $\mathbf{w}[j] = 1$ and

$$\Psi(\mathbf{K}[i], \mathbf{K}[j]) = \left(\frac{K_\ell[i] - K_\ell[j]}{r_{ij}} \right) \partial_\ell H(r_{ij}) \,.$$

Comparisons From Fig 4.1, we see that naive GPU implementations on `PyKeops` outperforms the `PBBFMM3D` implementation for $p < 5 \times 10^6$. Beyond this value, `PBBFMM3D` gets faster. It is likely that faster computations with the FMM would be obtained with a GPU implementation. Unfortunately, we did not find any robust and efficient GPU implementation of FMM.

Choice of step size

In our implementation, we use a combination of two step sizes. In the first 20 iterations, we use a fixed step size: $\eta^{(t)} = \eta$. As analyzed in [Cha+17], this strategy provides a convergence guarantee to a local minimizer of the cost function given that:

[2] see https://github.com/ruoxi-wang/PBBFMM3D.

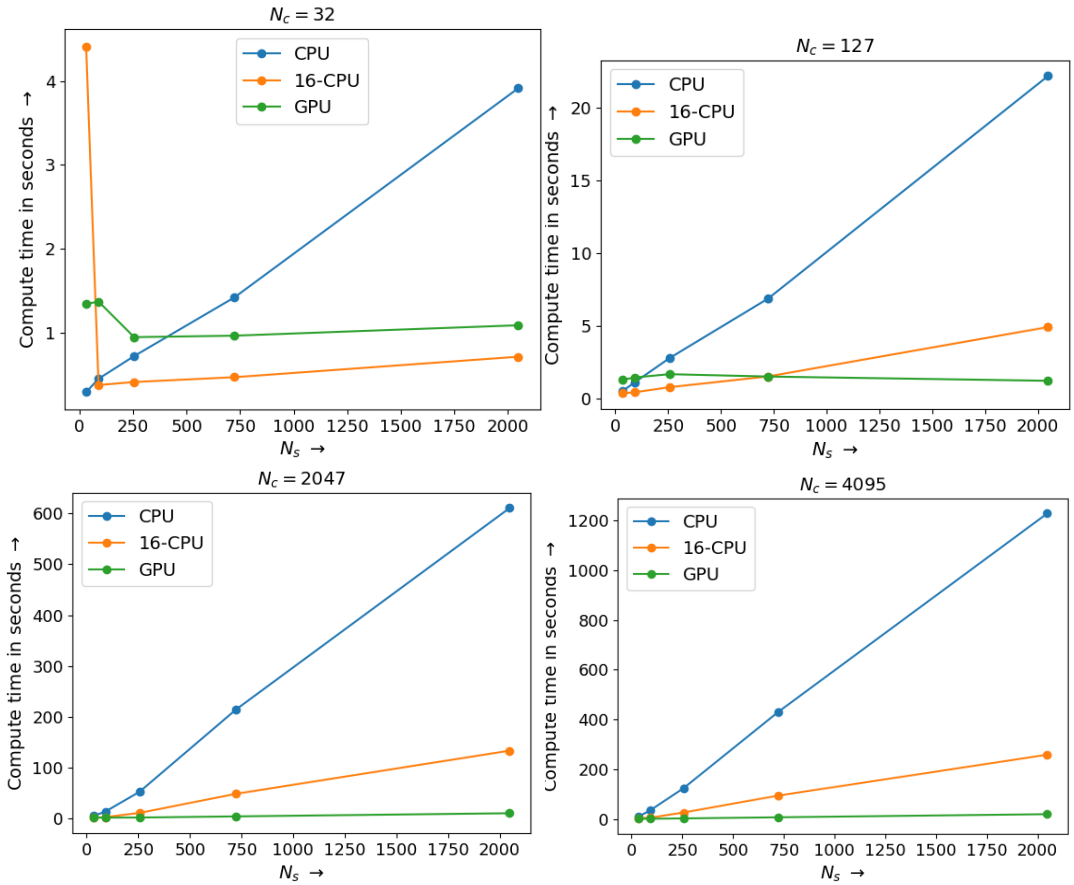


Figure 4.2 – Computation times for varying N_s and N_c for the projection step $\Pi_{\mathcal{Q}_{\alpha,\beta}^{N_c}}(\cdot, n_{pit})$ that was run over $n_{pit} = 200$ iterations with $G_{\max} = 40\text{mT/m}$ and $S_{\max} = 180\text{T/m/s}$.

- The kernel H has a L-Lipschitz continuous gradient.
- The step size is inversely proportional to the Lipschitz constant.

These conditions are satisfied with a regularized norm of the form $H(r) = \sqrt{r^2 + \epsilon^2}$. We can then set $\eta^{(t)}$ proportional to ϵ (i.e. 6.25^{-2}). The value of ϵ can be chosen as a fraction of the minimal distance between two points at a stationary point.

A constant step size is too conservative and a faster convergence can be obtained using a second-order dynamics close to the minimizer. This justifies switching to a Barzilai–Borwein [Bar+88] after first few iterations. Few theoretical guarantees are available for this technique, but it significantly accelerates the convergence empirically.

4.2.4 Projection step

The projection step in (4.6) for a general single k-space shot onto a given constraint set parameterized by (α, β) has been explored in [Cha+16; Cha+17]. For our implementations, we needed to extend the single shot iterative procedure called Algorithm 1 in [Cha+16] to projecting N_c shots. Note that the actual projection of a k-space shot is independent of other shots, and thus the computation can be done in parallel. Hence, we have implemented this step both on multi-CPU and GPU. To efficiently utilize a GPU, we used the CuPy module [Oku+17]. The computation times with different implementations for varying N_c and N_s are shown in Fig. 4.2. We found that the computation times vary linearly with N_s and are drastically reduced for the GPU implementation compared to the CPU versions (single and multicore). At lower N_s, we found that m-CPU and GPU implementations are latency

bound, giving anomalously higher computation times. However, the speedup obtained for larger N_s offsets these anomalous cases, giving an overall efficient implementation.

4.2.5 Multi-resolution strategy for faster convergence

In order to allow for the algorithm to reach faster convergence and lead to a better approximation of the target density, a multi-resolution approach as described in [Leb+19] was implemented. Under this methodology, the optimization of the sampling pattern was carried out on down-sampled curves. The interpolated solution was later used as a warm restart for the up-sampled problem. Our implementations involved dyadic scaling and up-scaling through simple linear interpolation of k-space shots. Let the linear interpolator be of the form $\mathcal{L}_{2d} : \Omega^d \to \Omega^{2d}$. We define the parameter N_d as the number of decimation steps in the algorithm. Note that the constraint space needs to be equally scaled with the problem, which results in scaling the α and β constraints mentioned in (7.2) to:

$$\alpha = \frac{\gamma G_{\max} 2^{N_d}}{2\pi K_{\max}}, \quad \beta = \frac{\gamma S_{\max} 2^{N_d}}{2\pi K_{\max}} \tag{4.13}$$

As we move through the dyadic decimation steps and up-sample the curve, these constraints are halved.

4.2.6 Overall algorithm

Algorithm Algorithm 2 summarizes how to concretely compute the SPARKLING solution along with multi-resolution steps described in Sec. 4.2.5. For more details on the iterative procedure involved in the projection step $\Pi_{\mathcal{Q}^{N_c}_{\alpha,\beta}}$, the reader can refer to [Cha+16].

Algorithm 2: Multi-resolution implementation of SPARKLING

Inputs : ρ, G_{max}, S_{max}, N_c, N_s, N_d, n_{git}, n_{pit}
Output: $\mathbf{K}$, the k-space sampling pattern

Initializations: $\mathbf{K}^{(0)} \in \Omega^{\frac{N_c \times N_s}{2^{N_d}}} = \Omega^p$
$\alpha \leftarrow \frac{\gamma G_{\max} 2^{N_d}}{2\pi K_{\max}}$, $\beta \leftarrow \frac{\gamma S_{\max} 2^{N_d}}{2\pi K_{\max}}$
while $N_d > 0$ **do**
  $p \leftarrow \frac{N_c \times N_s}{2^{N_d}}$
  for $t = 1 \dots n_{git}$ **do**
    $\mathbf{K}^{(t-1/2)} = \mathbf{K}^{(t-1)} - \eta^{(t)} \nabla F_p(\mathbf{K}^{(t-1)})$
    $\mathbf{K}^{(t)} = \Pi_{\mathcal{Q}^{N_c}_{\alpha,\beta}} \left(\mathbf{K}^{(t-1/2)}, n_{pit}\right)$
  // Warm restart next decimation step with linear interpolation
  // The dimension of k is doubled at each decimation step
  for $s = 1 \dots N_c$ **do**
    $\mathbf{k}_s^{(0)} \leftarrow \mathcal{L}_{\frac{N_s}{2^{N_d-1}}} \left(\mathbf{k}_s^{(n_{git})}\right)$
  $\mathbf{K}^{(0)} \leftarrow \left[\mathbf{k}_1^{(0)}, \dots, \mathbf{k}_{N_c}^{(0)}\right]$
  // Scale constraints
  $\alpha \leftarrow \frac{\alpha}{2}, \beta \leftarrow \frac{\beta}{2}$
  $N_d \leftarrow N_d - 1$

4.3 Numerical experiments and data acquisition

The sampling patterns were obtained by carrying out projected gradient descent as described above. With the above described improvements, the SPARKLING Generation time was just 10 minutes for 2D and nearly 6-9 hours for 3D on NVIDIA V100 with 5120 CUDA cores and 16GB DDR5X memory.

4.3.1 SPARKLING: a Python package

In the ethos of reproducible research and to move forward into better optimized patterns for MRI acquisition, all the implementations as described above is present in a Python package at the private repository[3]. All codes in the package scale to 2 and 3 dimensions directly, and most codes are agnostic and can be run on CPU or GPU with some change in run parameters. All the scanner constants and trajectory specification can be provided through a configuration file, and most of the codes are modular in nature. Interested researchers are requested to contact the authors for obtaining access to this package[4].

4.3.2 Acquisition parameters

With a goal of 600µm isotropic resolution in 3D MRI acquisitions, we planned to obtain a volume of $(N_x \times N_y \times N_z) = (384 \times 384 \times 208)$ size in order to cover the whole brain. For the sake of consistency, we used the same matrix size and resolution for our acquisitions on the NIST phantom[5]. The trajectories were generated for a clinical 3T MR system (Magnetom Prisma$^{\text{FIT}}$, Siemens Healthcare, Erlangen, Germany) with maximum gradient strength $G_{\max} = 40\text{mT/m}$ and peak slew rate $S_{\max} = 180\text{T/m/s}$. As the readout time was set to $T_{\text{obs}} = 20.48\text{ms}$ and the gradient raster time is $\Delta t = 10\mu\text{s}$, the number of samples per shot $\mathbf{k}_i$, was $N_s = 2048$. The number of shots N_c was varied based on the study described hereafter. For our in vivo studies, the k-space data was acquired on a Siemens 64 channel Head/Neck coil, while using 44 channels around the head during acquisition. The TE was 20ms and Repetition Time (TR) was 37ms. The flip angle was set to 15°, and the slice excitation was slab selective. We also obtained a reference volume collected using a 4-fold accelerated Cartesian acquisition (acquisition time or TA=15min 13sec) based on GRAPPA parallel imaging technique [Gri+02] with the same TE, TR and $T_{\text{obs}} = 20\text{ms}$. The projected gradient descent was carried out with multi-resolution decimation steps $N_d = 6$ for faster convergence. The algorithm was run for $n_{git} = 100$ outer gradient descent iterations with $n_{pit} = 100$ steps in the inner projection loop.

Choice of target sampling density

The target sampling density was chosen to be radially isotropic, which decays as an inverse polynomial with a constant plateau in the center of k-space. The density was defined with C, the cutoff frequency in k-space center having a constant density and D, the rate of decay for higher frequencies. Mathematically, we define the target density $\pi_{C,D}(x) : \Omega^d \to \mathbb{R}_+$ as follows:

$$\pi_{C,D}(x) = \begin{cases} \kappa & |x| < C \\ \kappa \left(\frac{C}{|x|}\right)^D & |x| > C \end{cases} \tag{4.14}$$

where κ is a constant obtained through normalization as $\kappa = \frac{1-D}{2C(C^{D-1}-D)}$. The resulting density is radially symmetric and is of the form described in Fig. 4.3. The choice of a radial density was motivated by the wish to provide rotation invariant reconstruction results. Notice that the recent learning based approaches [She+20; Bah+20b] result in non-symmetric densities. This is probably due to the fact that brains or knees databases such as fastMRI [Zbo+18] used for training have boundaries which are dominantly vertical or horizontal. However, the fine details may be in arbitrary orientations.

The choice of the density was carried out by grid searching for optimal parameters $\hat{C}$ and $\hat{D}$ on the target sampling distribution as defined in (B.1). We performed retrospective reconstruction on complex Cartesian reference (with phase to account for off-resonance artifacts by phase accrual) in vivo brain data obtained through virtual coil combination[Par+14]

[3] https://gitlab.com/cea-cosmic/CSMRI_sparkling
[4] It cannot be made open source given patent application.
[5] NIST Phantom

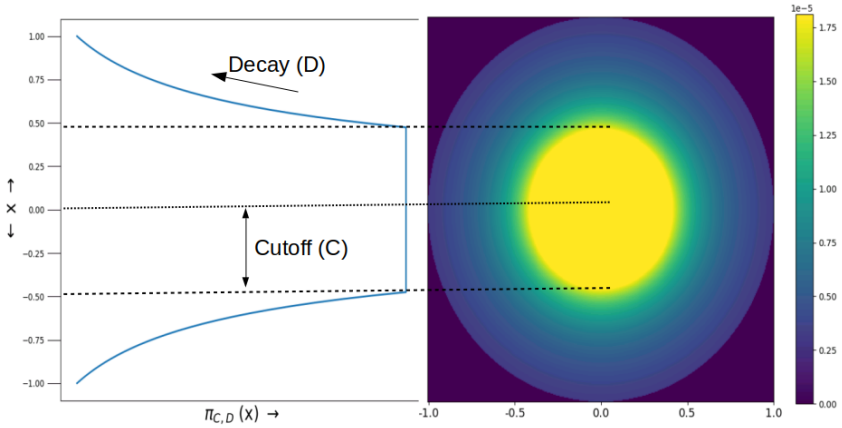


Figure 4.3 – Parameterization of variable density with cutoff C and decay D.

of raw multi-channel images (see Sec. A.3 in Appendix). We chose a target sampling distribution of $\hat{C} = 25\%$ and $\hat{D} = 2$ as it is gives the best image quality as well as more reliable image reconstruction performance in the single-channel setting. Improved reconstruction performances can be observed using an optimal target density of $\hat{C} = 1\%$ and $\hat{D} = 1$ for our multi-channel coil configuration (see Fig. A-4). However, this setting was not retained in this work to ensure that optimized 3D SPARKLING trajectories are generic and do not specifically depend on our given coil configuration (Siemens 64Rx head/neck).

Initialization and Perturbation

As the problem being solved in (5.3) is non-convex, different choices of initialization would lead to different solutions. In [Laz+19] for 2D imaging we observed that radial initialization performed the best for exploring the k-space. Hence, here for 3D imaging we stick to 3D radial initialization too. For the sake of simplicity, and also to ensure radially symmetric initialization, we set up the trajectories with $\sqrt{N_c}$ shots in x-y plane and then rotate each shot $\sqrt{N_c}$ times along an in-plane axis orthogonal to the shot. More generic solution can be obtained by solving for the minimum electrostatic potential energy configuration of N_c electrons over the surface of a unit sphere, however this approach was not pursued in this work.

For best reconstructed image results, we would want each k-space shot to maximally explore the k-space. The 3D radial initialization is too structured with each k-space shot traveling only from end to end of k-space. To enable a broader k-space exploration and obtain a better minimizer of the original problem, we added a perturbation to each initial shot. To achieve this, we perturbed each trajectory sample point in k-space by adding zero mean uniform random noise along each dimension. Particularly, we compared the resulting optimized trajectory obtained after a perturbation as a random motion of each k-space point with maximum amplitude set at 0.1 and 0.75 (we remind that the sampling domain is normalized to $\Omega = [-1, 1]^3$). The optimized trajectory patterns are presented in Fig. 4.4. We clearly show that with more perturbation, the k-space trajectory tends to explore a broader part of k-space giving a better coverage overall. We also notice quantitatively that with more perturbation the value of the cost function converges to a lower local minimum. Further, we would like to emphasize that these trajectories are particularly useful in cases of high receiver sampling rates, as they would then sample more of the k-space per shot and would overall prevent the presence of any hole in the sampling pattern.

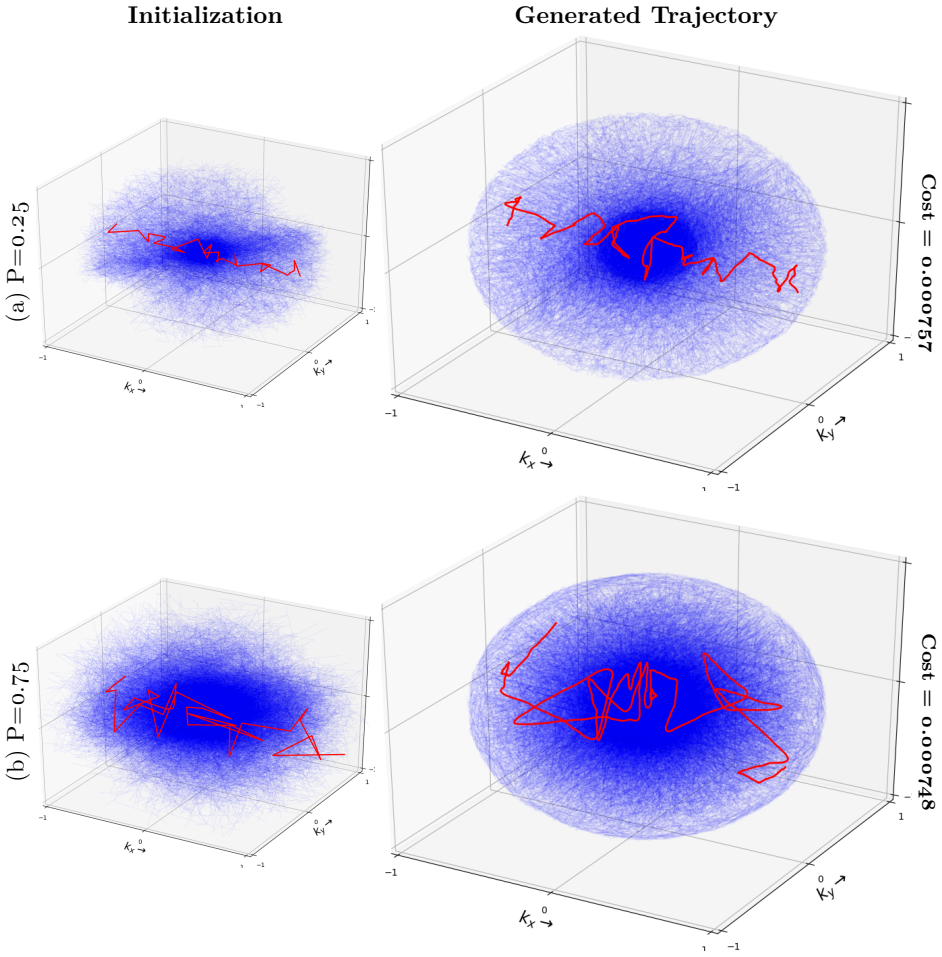


Figure 4.4 – Effect of adding a perturbation (P) to the initial k-space trajectory in $\Omega = [-1, 1]^3$ as zero mean uniform random noise at each trajectory sample. Trajectories are generated with maximum displacement of k-space point to (a) 0.25 and (b) 0.75 in the initialization. The left side of the figure is the initialization to SPARKLING algorithm and the right is the output of the algorithm. We also present the values of the cost obtained with (5.3).

4.4 Results

4.4.1 From trajectories to k-space data

The k-space data $\mathbf{Y} = (\mathbf{y}_i)_{i=1}^{N_c}$ is sampled by the ADC at the dwell time period δt. In practice, the dwell-time δt is a fraction of the raster time Δt and was set to $\delta t = 2\mu$s. This means that $\mathbf{y}_i \in \mathbb{C}^m$ with $m = N_s \left\lfloor \frac{\Delta t}{\delta t} \right\rfloor$ the number of measurements per shot. Overall, we collect $M = N_c\, m$ k-space data points in $\mathbf{Y}$. Consequently, during the image reconstruction process, we obtain the k-space locations of $\mathbf{Y}$ by linearly interpolating the optimized trajectory $\widehat{\mathbf{K}}$ originally sampled at Δt, to the δt period.

4.4.2 Assessment of point spread function

We present the full 3D SPARKLING , obtained with $N_c = 4096$ in Fig. 4.5. We visualize the trajectory along the mid-planes of 3 orthogonal orientations and provide an approximate sampling mask in these planes. Further, to understand why these trajectories are expected to yield good image reconstructions, we measure the 3D PSF. Each point spread function was computed by taking a density compensated Nonuniform Fast Fourier Transform (NUFFT) adjoint of k-space measurements set to 1 ($y_i[n] = 1, \forall i = 1, \cdots, N_c, \forall n = 0, \cdots, m-1$)

as described in [Pau]. In Fig. 4.6 we compare the PSF with respect to earlier generated spherical stack of 2D SPARKLING (SpSOS) trajectories. Particularly, we emphasize the reduction in sidelobes along the z axis. Further, for the purpose of numerical comparison, we computed the Full Width at Half Maximum (FWHM), Peak-to-Sidelobe Level (PSL) and Peak-to-Noise Level (PNL) in Tab. 4.1. The FWHM is calculated as the width of the peak of PSF at half of the maximum value and the PSL and PNL are calculated in dB as presented in Fig. 4.6(d).

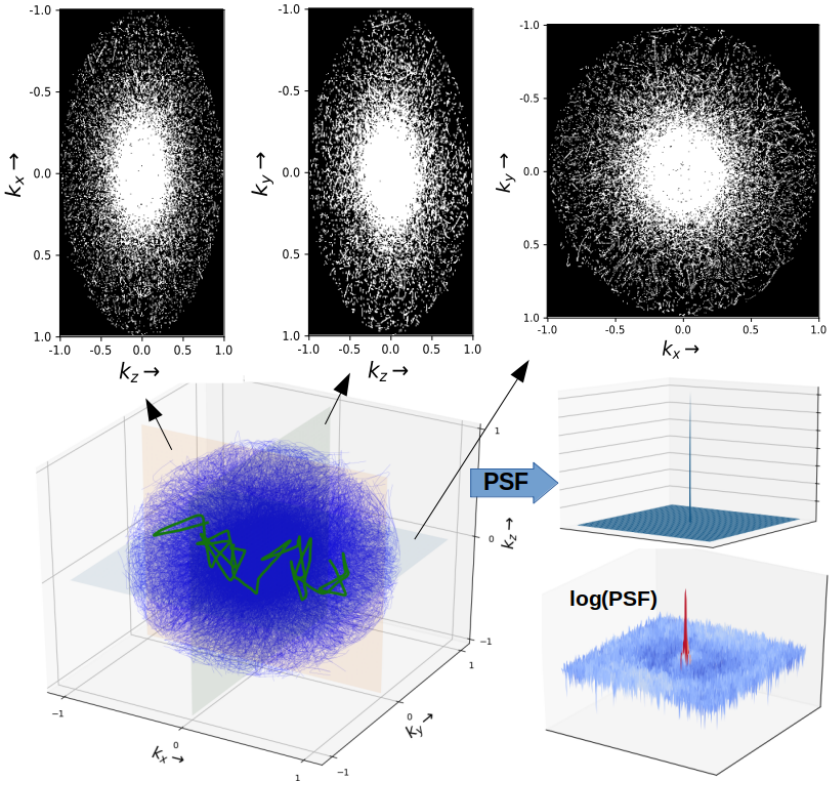
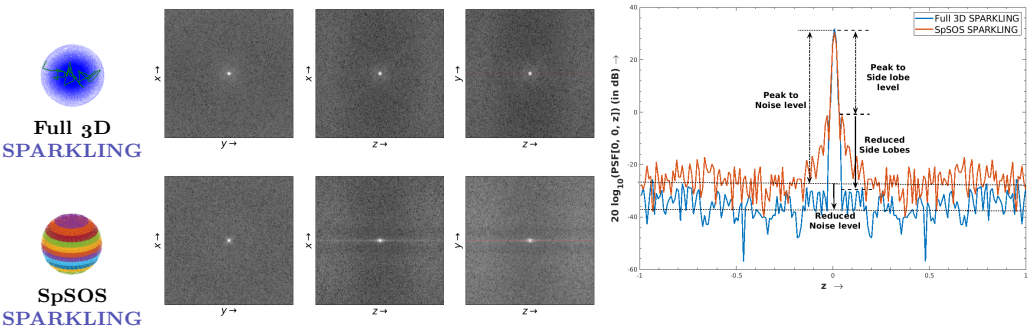


Figure 4.5 – Full 3D SPARKLING Trajectory for $N_c = 4096, N_s = 2048$ and the PSF along the mid z-plane computed from the sampling mask (measurements sampled at the dwell-time period δt).

Table 4.1 – Comparing metrics of PSF with FWHM (lower is better), PSL and PNL (higher is better).

Trajectory	FWHM (in voxel units)			PSL (in dB)	PNL (in dB)
	x	y	z		
Full 3D	2.3	2.4	2.5	**35.80**	**67.25**
SpSOS	2.4	2.4	2.5	31.65	65.44

As shown in Fig. 4.6, we see that the full 3D pattern provides us with much higher PSL (4.15 dB more) and PNL (1.81 dB more), two quantitative indices that demonstrate the full 3D SPARKLING methodology outperforms the spherically stacked version. In contrast, we observe that the FWHM is nearly the same for both methods, even though the FWHM along the x axis is slightly lower for the full 3D pattern. However, this minor difference in FWHMs and the slight anisotropy in FWHM can be explained by the fact that the full 3D initialization was severely perturbed (0.75) as described in Sec. 4.3.2.



(a) **(b)** **(c)** **(d)**

Figure 4.6 – Comparison of PSF between full 3D SPARKLING and SpSOS sampling masks (measurements collected at the dwell-time period over the corresponding trajectories). The logarithm of 3D PSF (in voxel units) are viewed along the mid-slices in **(a)** axial plane (x, y, 0), **(b)** sagittal plane (x, 0, z) and **(c)** coronal plane (0, y, z). **(d)** The PSF are compared in logarithmic scale along the z direction.

4.4.3 Non-Cartesian MR image reconstruction

All MR images that rely on non-Cartesian k-space data in this paper were reconstructed using a self-calibrated synthesis-based CS reconstruction algorithm [PM99; Kno+14; Gue+20; El +18a] whose details are provided in Appendix (cf. Sec. A.2). For the sake of reproducibility, the code for MR image reconstruction is made open source in `pysap-mri`[6], a plugin of the PySAP software [Far+20a]. Of course, future work will combine deep-learning based image reconstruction with full 3D SPARKLING .

In this work, we did not carry out off-resonance artifact corrections using [Sut+03], as it is beyond the scope of this manuscript. However, note that this does not require any supplementary scan for obtaining $\Delta B0$ map as the latter can be directly estimated from phase information using [Dav+21]. For the sake of completeness, we show in appendix the performance of our trajectory with off-resonance corrections for Acceleration Factor (AF)=10 in Fig. A-7.

4.4.4 Phantom

Retrospective studies

We proceed by carrying out a retrospective study to assess the quality of reconstructed images. We varied the AF $= \frac{N_y \times N_z}{N_c}$ for 3D MR imaging, i.e. computed with respect to fully sampled data) from 10 (TA=4min 58sec) to 40 (TA=1min 16sec) compared to a fully sampled scenario or equivalently from 2.5 to 10 compared to the reference Cartesian p4 (i.e. AF=4) under-sampled acquisition, reconstructed using the GRAPPA algorithm [Gri+02]. Our motivation was to understand the degradation in image quality while decreasing the number of collected spokes. Further, a study was also carried out with the TPI [Boa+97], as a comparison with a non-Cartesian reference from the literature. The results are presented in Tab. 4.2. They clearly show that the optimized full 3D SPARKLING strategy is robust to high acceleration factors in terms of image quality as reflected by the higher SSIM scores. In contrast, the performances of the SpSOS approach start to get worse already for AF=20. Finally, the SSIM score for TPI for AF=10 is already significantly lower than that of SpSOS. The reconstructed images are presented in the Appendix (see Fig. A-5).

[6] https://github.com/CEA-COSMIC/pysap-mri

Table 4.2 – Comparing SSIM metrics of retrospective phantom image reconstruction.

Trajectory	AF10	AF15	AF20	AF40
Full 3D	**0.964**	**0.935**	**0.923**	**0.816**
SpSOS	0.927	0.864	0.737	0.575
TPI	0.63	0.592	0.573	-

Prospective acquisition

In order to understand how the effective spatial resolution compares to the target resolution (here 0.6mm isotropic), we performed prospective acquisitions on the NIST phantom for full 3D SPARKLING and SpSOS trajectories with varying AF (AF=15 and 20 for both, AF=40 for full 3D). The results are presented in Fig. 4.7. Particularly, we show a slice that includes the resolution insets present on the NIST phantom (coffin of plate 4). The latter can be used to estimate the effective resolution. This slice consists of 5 resolution insets, each having 2x16 circles. The diameters of these circles vary linearly from 0.8mm down to 0.4mm. The inter-circle space (measured between the centers of the circles) also reduces linearly from 1.6mm down to 0.8mm in steps of 0.2mm.

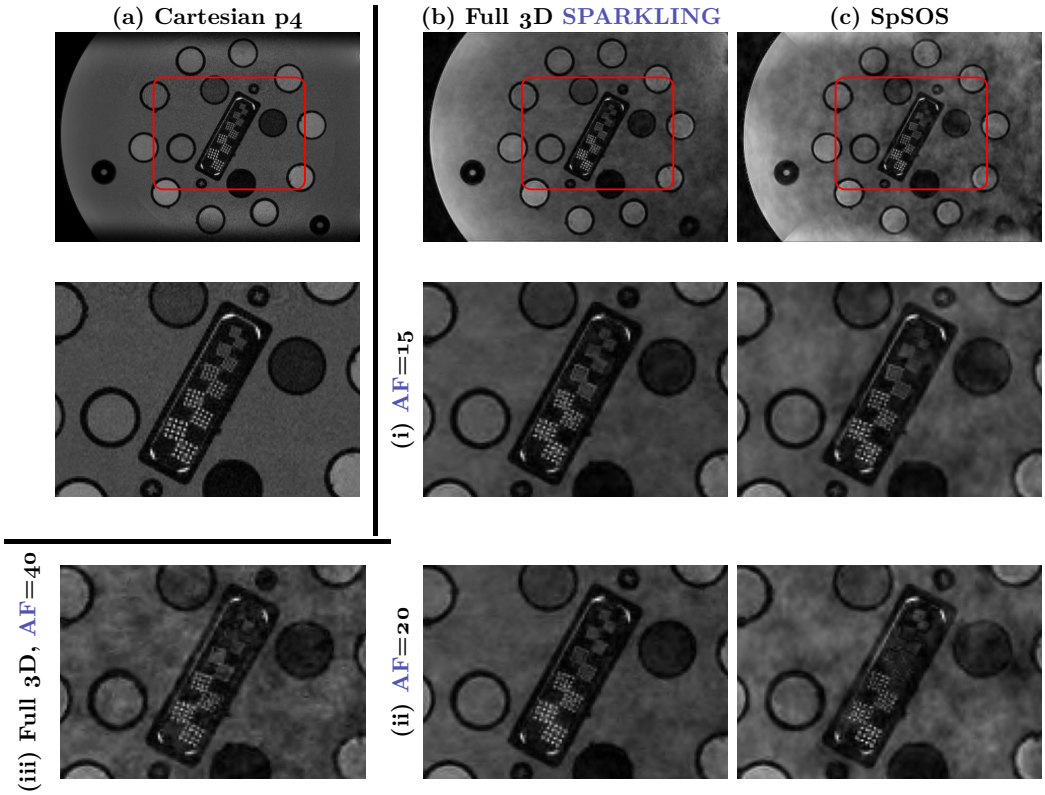


Figure 4.7 – Comparison of resolution insets for Full 3D SPARKLING (b) and SpSOS (c) with prospective phantom scans at (i) AF=15, (ii) AF=20 as compared to Cartesian p4 (a). Additionally, we present the results for full 3D SPARKLING trajectory at (iii) AF=40 at the bottom-left.

We see that the intensity profile of our reconstructed MR images does not follow that of Cartesian reference as we did not carry out coil sensitivity normalization in our reconstructions. This can be performed using the rapid pre-scan coil sensitivity measurements done in a few seconds. This point will be addressed in future works. Overall, we observe that full 3D SPARKLING trajectories provide less noisy images compared to SpSOS ones. Further, it is worth noting that at AF=10 and 20, we can distinguish in between the resolution insets

down to 0.7mm, with an increasing noise level over the image for a higher acceleration factor. However, we observe some resolution loss over the images based on SPARKLING trajectories (more pronounced for SpSOS), where we see some blurring for circles of diameter 0.6mm separated by 1.2mm (taken from the center of circle). This helps us understand the expected degradation in image resolution. Therefore, the effective image resolution is estimated to be 0.6-0.7mm isotropic at AF=10 and 20 and is evolving toward 0.7-0.8mm at AF=40 for full 3D SPARKLING .

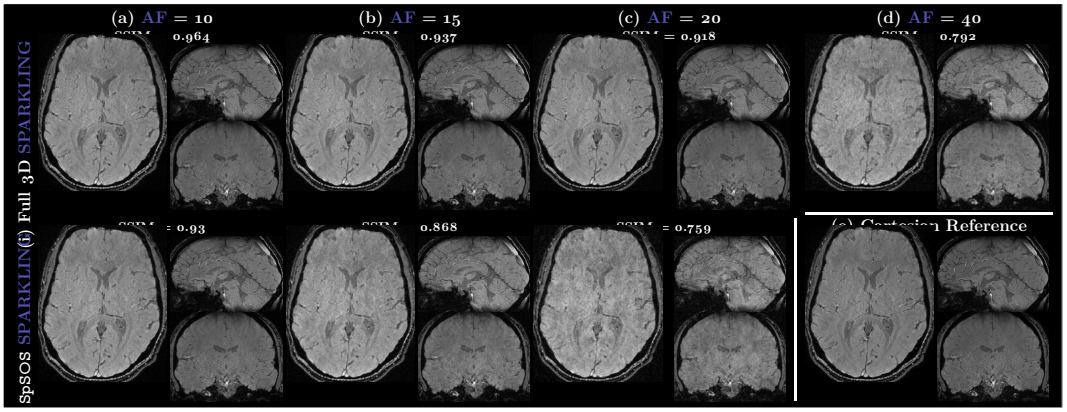


Figure 4.8 – Comparison of retrospective results for (i) fully optimized 3D SPARKLING (top row) and (ii) SpSOS (bottom row) for varying acceleration factors (from left to right, AF=10 (a), 15 (b) and 20 (c)) on in vivo human brain scans. Cartesian Reference (e) is provided for comparison and results for full 3D trajectory at AF 40 (d) is also presented. SSIM scores are reported for each setup.

4.4.5 In vivo

We collected in vivo data with full 3D SPARKLING and SpSOS trajectories for brain imaging on a healthy volunteer (male, 25 y.o.). This study was approved by a national ethics committee (CPP 100048). The volunteer signed a written informed consent form.

Retrospective studies

To understand how the trajectories perform for in vivo brain data, we repeat the earlier retrospective study on Cartesian p4 scans acquired on the volunteer. The results of the scans are presented in Fig. 4.8. We see that full 3D SPARKLING trajectories outperforms the SpSOS trajectories both visually and quantitatively in SSIM metrics with maximum SSIM of 0.964 (AF=10). Moreover we show that both SPARKLING trajectories outperform TPI in Fig. A-6 in Appendix. Additionally, the SSIM metrics follow the similar trend as seen for phantom data, with the SSIMs for SpSOS dropping off more rapidly from 0.93 (AF=10) to 0.759 (AF=20). In contrast, the full 3D SPARKLING trajectories tend to preserve the structures (SSIM scores above 0.9 at AF=15 and 20) and show some blurring artifacts only at AF=40 where SSIM drops to 0.792. Particularly, it is interesting to note that full 3D SPARKLING at AF=40 outperforms SpSOS at AF=20.

Prospective acquisition

Finally, we collected prospectively accelerated in vivo data at 3T on the same individual using the same SPARKLING trajectories. We present the reconstructed images for various accelerations factors in Fig. 4.9. They clearly show superiority of the full 3D SPARKLING pattern compared to SpSOS. Image quality is well preserved for AF=10 and 15 and slightly noisy at AF=20 in full 3D strategy (Fig. 4.9, top row), while it tends to get noisy at AF=15 and severely impaired at AF=20 for SpSOS strategy (Fig. 4.9, bottom row). Moreover, we

observed that the quality of AF=15, AF=20 and AF=40 in full 3D strategy is comparable to AF=10, AF=15 and AF=20 in SpSOS pattern respectively, allowing for an additional 2x shorter scan time. Further, we found that the full 3D SPARKLING pattern at AF=15 is comparable to GRAPPA Cartesian p4.

It is important to note that the volunteer slightly moved between some scans, hence prospective image comparisons can only be carried out qualitatively. To better understand reconstruction quality, we present zoomed in visualizations for prospective result at AF=10 in Fig. 4.10. Further, for the sake of comparison between retrospective simulations and actual prospective scans, we also show the retrospective results for AF=10 with full 3D trajectory. We find that full 3D strategy retains better structures of the brain in the MR image than SpSOS, which is clearly visible in the cerebellum in the sagittal view.

The comparison with retrospective image allows us to directly identify some degradation and loss of small details in prospective images. Potential explanations for this effect are the T_2* blurring and off-resonance artifacts, which drastically drop the effective SNR obtained (see Fig. A-2 in Appendix). This confirms that in vivo acquisitions are more challenging.

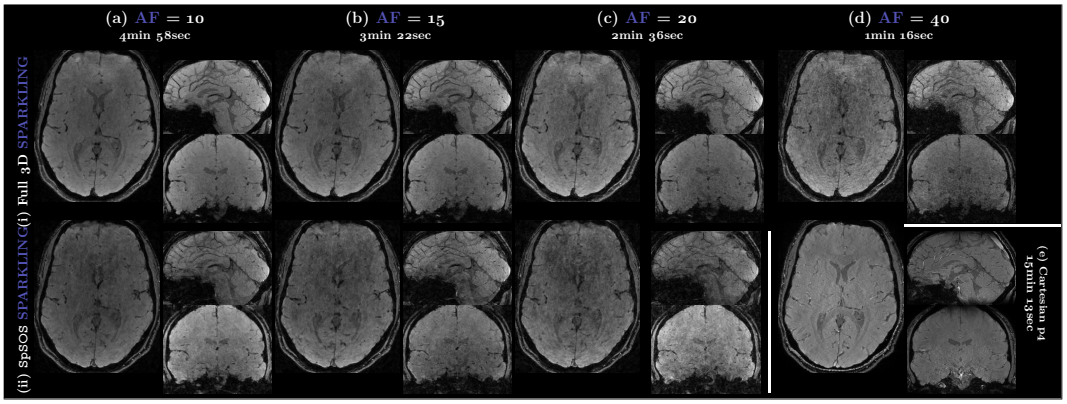


Figure 4.9 – Comparison of prospective results for (i) fully optimized 3D SPARKLING (top row) and (ii) SpSOS (bottom row) for varying acceleration factors (from left to right, AF=10 (a), 15 (b) and 20 (c)) on in vivo human brain scans. Cartesian p4 scan (e) is provided for comparison and results for full 3D trajectory at AF 40 (d) is also presented. The scan times are reported for each AF.

4.5 Discussion

One key aspect of optimized full 3D SPARKLING trajectories is that it results in a sampling pattern that enforces variable density sampling in all the 3 dimensions. We hypothesized that this allows us to efficiently under-sample the k-space acquisitions, thus making it possible to push the acceleration factor to a larger value than what was achieved earlier, while still maintaining a good image quality. The current work actually demonstrates that at fixed acceleration factor, full 3D SPARKLING significantly outperforms the stacking strategy [Laz+20a] in terms of image quality. Alternatively, we show that this gain can be translated into shorter scan time by a factor of one third (AF=15 for full 3D vs AF=10 for SpSOS) for a given image quality.

Further, the full 3D trajectory is constrained to pass through the center of k-space for each shot at echo time. This ensures that we obtain the lower frequency image content repeatedly, hence we can potentially use these trajectories for motion correction. Also, as the center of k-space is visited repeatedly at different time intervals in scan, this allows for easy adaptability of this trajectory for dynamic imaging like functional MRI. Such a trajectory can also be used for correcting certain artifacts causing off-resonance effects, which are due to static and dynamic B0 inhomogeneities (heart beat, breathing). A preliminary solution has been proposed for static B0 inhomogeneities estimation and correction in [Dav+21].

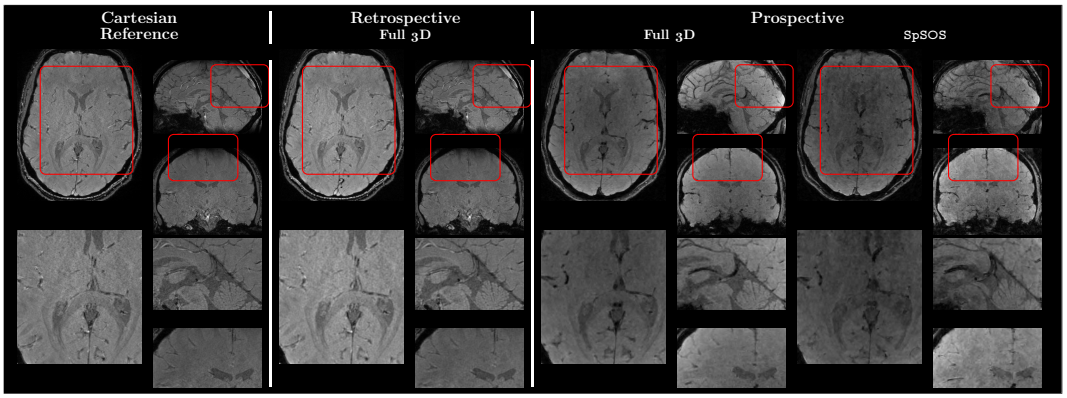


Figure 4.10 – Comparison of prospective results for fully optimized 3D SPARKLING (right-left) and SpSOS (right most) with Cartesian reference (left) and retrospective full 3D SPARKLING (center) for AF=10 (scan time = 4min 58sec). In each panel of the top row, axial (left), sagittal (top right) and coronal (bottom right) slices are shown and a red frame is delineated in the central part of the brain for zooming purpose. Bottom row shows the magnified views with the same layout (axial, sagittal and coronal slices in the left, top-right and bottom right insets, respectively).

As the developed trajectories exploit the scanner hardware constraints nearly to the maximum, it is worth paying attention to the eddy current effects on the trajectory. To this end, we measured the trajectory with the help of a Skope field camera [De +08] and observed in Sec. A.1.3 that the error between the prescribed and actual trajectories is minimal (cf. Fig. A-3).

While the current SPARKLING algorithm is generic and can be applied to any imaging contrast a priori, we choose T_2^*-w imaging as it allows us to keep longer T_{obs} hence enabling a full exploration of 3D k-space. In order to understand the effects of T_2^* blurring and off-resonance, we simulated the PSF under these scenarios in Sec. A.1.2. Additionally, as we interfaced a GRE pulse sequence (FLASH in the Siemens taxonomy) with the SPARKLING outputs, the adaptation of this algorithm to other contrasts (e.g. T_2) would potentially need the development of a turbo spin echo (TSE) sequence that is able to play arbitrary gradients. These developments are left for future work. However, prospective implementations of T_1-w contrast have already been done in 2D for comparison with BJORK in [Wan+21] (outside of our group).

One limitation of SPARKLING is that the original optimization problem (5.3) is non-convex and the fact we used a locally convergent optimization algorithm to compute a minimizer. Hence the final solution heavily depends on its initialization. To overcome this issue, we introduced some perturbation (uniform random noise in the k-space locations) and illustrated in Fig. 4.4 that a larger perturbation results in a much better k-space coverage, allowing us to reach a better minimizer to the original optimization problem. However, there is no theoretical guarantee this approach provides a systematic better solution as the underlying optimization process remains rather disconnected from MR image reconstruction and the maximization of image quality.

In the same vein, another limitation of the resulting reconstructed MR images is that they heavily depend on the target sampling distribution. We obtained our results by parameterizing this distribution, thereby optimizing for its parameters using a grid search on in vivo brain data. However, these optimal parameters are not generalizable for different contrast and organs. Further, such parametrization can prevent us from using more complex target sampling densities. To overcome this limitation, ongoing work intends to couple SPARKLING with the learning of the target sampling density from the magnitude spectrum of human brain MR images [CRC21]. Akin to this work, we could also jointly optimize for the acquisition (sampling pattern) and reconstruction schemes (regularization parameters) under MR hardware and imaging contrast constraints, either in a bilevel optimization [She+20]

or using deep learning approaches [Wei+20; Ved+20; Wan+21; CRC22]. These extensions would help us to take some factors, like the anatomy and the imaging contrast, into account in the design of trajectories with perfectly matched target sampling densities for these cases.

4.6 Conclusion

In this paper, we proposed an optimization for full 3D SPARKLING k-space trajectories for accelerated high resolution 3D magnetic resonance imaging and demonstrated its superiority over the previously proposed stacking strategies on phantom and in vivo human brain data at 3T for the particular case of T_2*-weighted imaging. We discussed the major computational bottlenecks that prevented us earlier from proceeding towards these full 3D trajectories. We then derived some implementations (GPU and multi-CPU) that helped us massively accelerate the original algorithm. Our results showed that a 600µm isotropic scan on human brain is achievable in 1min 16sec, whereas 3m 22sec is required to reach image quality comparable to GRAPPA-4 parallel imaging. Overall, this is a significant step forward for CS acquisitions in MRI. Future work will be devoted to the extension to 4D imaging, namely for fMRI.

* * *
* *
*

Chapter 5

Reducing artifacts in SPARKLING

This chapter covers content that has been submitted to a peer-reviewed journal:

- **Chaithya, G R**, G. Daval-Frérot, A. Massire, A. Vignaud and P. Ciuciu. "Improving SPARKLING trajectories through Minimized Off-Resonance Effects and Gridding of Low Frequencies". In: *under review MRM*

A part of this work was also presented in an international peer-reviewed conference with proceedings:

- **Chaithya, G R**, G. Daval-Frérot, A. Massire, B. Mailhe, M. Nadar, A. Vignaud and P. Ciuciu. "MORE-SPARKLING: Non-Cartesian trajectories with Minimized Off-Resonance Effects". In: *ISMRM*. 1435. London, UK, May 2022

5.1 Introduction

Non-Cartesian (NC) sampling trajectories are crucial to have optimal k-space coverage, to help reduce the acquisition times in Magnetic Resonance Imaging (MRI). This reduction in scan time is instrumental in increasing patient throughput and reducing the motion artifacts

as the patient spends a shorter time period in the scanner environment. In this regard a lot of parameterized NC trajectories have been explored in the literature like radials, spirals and rosette [Ahn+86a; Mey+92; Jac+92; Nol97; Law+09; Lus+05a]. However, based on CS theories, efficient ways to undersample the k-space is through VDS, where the center of k-space (low frequencies) is sampled more densely than its periphery (higher frequencies). While the traditional NC trajectories do sample the k-space using VDS, they do not enforce a user-defined TSD in the k-space, which could be crucial for tailoring optimum sampling strategies based on a given target organ, contrast and coil configuration.

Recently, SPARKLING [Laz+19] was introduced as a means to optimize k-space sampling pattern according to a prescribed TSD while each underlying NC k-space trajectory followed the MR hardware constraints, particularly maximum gradient G_{max} and slew rate S_{max}. This work was successfully extended to 3 dimensions [Cha+22d] which enabled reduction in acquisition times (nearly 4× as compared to GRAPPA 4 [Gri+02]) with negligible degradation in retrospective reconstructed image quality.

However, a large gap was observed in the reconstructed image quality between simple retrospective simulations and actual prospective scans (see [Cha+22d, Fig. 10]) in the case of T_2^*-w imaging. Such a discrepancy in prospective setting was identified [Cha+22d, S2.B],[Dav+22] to be due to the presence of strong off-resonance effects and T_2^*decay whose impact is accumulated in T_2^*-w contrast due to longer TE used to enhance the susceptibility contribution. This effect is amplified in NC imaging and notably in 3D SPARKLING as such trajectories have arbitrary readout directions, leading to local k-space inconsistencies. Although these artifacts can be corrected [Sut+03] without needing any supplementary scan for ΔB_0 map [Dav+22], such corrections are computationally expensive. In this work, we insert temporal weights into the cost function of the SPARKLING algorithm, giving us temporally smooth k-space trajectories which present with MORE.

Additionally, another limitation of the SPARKLING trajectories is that we use affine TE constraints, where we limit the shots to pass through the center of k-space at echo time to obtain images at chosen target contrast. This results in strong oversampling of the center of k-space with respect to the Nyquist criteria which can be detrimental to image quality as it results in increased ΔB_0 artifacts due to multiple trajectories crossing the center of k-space along different trajectory paths. Further, such oversampling is sub-optimal as these extra samples can be used to sample higher frequencies resulting in improved image reconstructions with finer details in structures. Although this is counterintuitive to CS theories which justifies VDS theoretically, note that CS theories are relative and do not prescribe how many samples must be collected in practice. Having a TSD which enforces k-space samples beyond the Nyquist criteria can lead to suboptimality. We tackle this issue by updating the constraint set in SPARKLING algorithm giving us trajectories with gridded sampling at center of k-space.

With this, we introduce novel MORE and GoLF features which can be used individually and in conjunction to design trajectories that exhibit minimized ΔB_0 artifacts and improved reconstructed image quality. We limit ourselves to the case of T_2^*-w imaging where maximal ΔB_0 artifacts are accumulated. Note that while this work is specifically applied to the design of SPARKLING trajectories, the constraints and penalties developed are more generic and could be used more widely in any trajectory optimization process including state-of-the-art learning based NC trajectory design like PILOT [Wei+20; Ved+20], BJORK [Wan+21] and HybLearn [CRC22].

The sections below are organized as follows. We first develop the required theory for the case of 3D non-Cartesian imaging in Sec. 5.2, then we briefly discuss the limitations and ways to tackle them in the current algorithm in Sec. 5.3. Later in Sec. 5.4, we proceed to fine tune our algorithm through PSF analysis and grid-search on prospectively acquired k-space data on 3T. Finally, we carry out benchmark studies on both *in silico* and *in vivo* settings prospectively to demonstrate the gain in scan acceleration obtained by using these novel improvements.

5.2 Theory

Following the formulation in [Cha+22d], the k-space domain that is acquired for a 3D MR volume of size $N_x \times N_y \times N_z$ over a field-of-view (FOV) $\text{FOV}_x \times \text{FOV}_y \times \text{FOV}_z$, is defined within $[-K^x_{\max}, K^x_{\max}] \times [-K^y_{\max}, K^y_{\max}] \times [-K^z_{\max}, K^z_{\max}]$, with $K^\ell_{\max} = \frac{N_\ell}{2FOV_\ell}$ and $\ell = x, y, z$. For the sake of simplicity, in what follows we assume the same spatial resolution and FOV along the three dimensions so $K^x_{\max} = K^y_{\max} = K^z_{\max} = K_{\max}$ and $\text{FOV}_x = \text{FOV}_y = \text{FOV}_z = \text{FOV}$. Also, the 3D k-space sampling domain is normalized to $\Omega = [-1, 1]^3$.

We optimize the 3D k-space sampling pattern $\mathbf{K} = (\mathbf{k}_i)_{i=1}^{N_c}$ which is composed of N_c shots, each 3D shot $\mathbf{k}_i(t) = (k_{i,x}(t), k_{i,y}(t), k_{i,z}(t))$, being controlled by magnetic field gradients $\mathbf{G}_i(t) = (G_{i,x}(t), G_{i,y}(t), G_{i,z}(t))$ as follows: $\mathbf{k}_i(t) = \frac{\gamma}{2\pi} \int_0^t \mathbf{G}_i(\tau) d\tau$, with γ the gyro-magnetic ratio ($\gamma = 42.57$MHz/T for proton imaging). In contrast to 2D imaging, here in 3D each $\mathbf{G}_i(t)$ is played by the scanner throughout the readout duration T_{obs} at a pace of gradient raster time (Δt) resulting in a number of samples per shot $N_s = \left\lfloor \frac{T_{\text{obs}}}{\Delta t} \right\rfloor$. The k-space data is later sampled at the Analog to Digital Converter (ADC) at every dwell time δt which is a fraction of Δt.

5.2.1 Trajectory Constraints

Hardware constraints on the maximum gradient amplitude ($G_{\max}$) and slew rate ($S_{\max}$) induce limitations in trajectory speed and acceleration, respectively. These limits can be expressed as box constraints on the amplitude of the discrete derivatives of the k-space trajectory $(\mathbf{k}_i[n])_{n=1}^{N_s}$, where $\mathbf{k}_i[n]$ is the obtained by discretizing $\mathbf{k}(t)$ at $n\Delta t$, with Δt the gradient raster time. We obtain these constraints from [Cha+22d, Eq. (2)] as:

$$\mathcal{Q}^{N_c}_{\mathbf{A},\underline{\mathbf{b}}} = \left\{ \begin{array}{c} \forall i = \{1, \dots, N_c\}, \quad \mathbf{k}_i \in \Omega^{N_s}, \\ \mathbf{A}_i \mathbf{k}_i = \underline{\mathbf{b}}_i, \\ \|\mathbf{k}_i\|_\infty \leq 1, \ \|\dot{\mathbf{k}}_i\|_{2,\infty} \leq \alpha, \ \|\ddot{\mathbf{k}}_i\|_{2,\infty} \leq \beta, \end{array} \right\} \tag{5.1}$$

where

$$\begin{aligned} \dot{\mathbf{k}}_i[n] &= \frac{\mathbf{k}_i[n] - \mathbf{k}_i[n-1]}{\Delta t} \\ \ddot{\mathbf{k}}_i[n] &= \frac{\mathbf{k}_i[n+1] - 2\mathbf{k}_i[n] + \mathbf{k}_i[n-1]}{\Delta t^2} \\ \|\mathbf{c}\|_{2,\infty} &= \sup_{0 \leq n \leq N_s - 1} \left(|c^x[n]|^2 + |c^y[n]|^2 + |c^z[n]|^2 \right)^{1/2}, \end{aligned}$$

for all $\mathbf{c} \in \Omega^{N_s}$ and (α, β) are obtained by normalizing hardware and Nyquist constraints to the sampling domain Ω (see [Cha+22d, Eq. (2a-b)]).

The purpose of $\mathbf{A}_i = (\mathbf{a}_{i,1} | \mathbf{a}_{i,2} | \cdots | \mathbf{a}_{i,c_i})^T \in \{0,1\}^{3c_i \times N_s}$ with

$$\mathbf{a}_{i,j} = \begin{bmatrix} \mathbf{a}^x_{i,j}[1] & \mathbf{a}^y_{i,j}[1] & \mathbf{a}^z_{i,j}[1] \\ \vdots & \vdots & \vdots \\ \mathbf{a}^x_{i,j}[N_s] & \mathbf{a}^y_{i,j}[N_s] & \mathbf{a}^z_{i,j}[N_s] \end{bmatrix}$$

and $\underline{\mathbf{b}}_i = (\mathbf{b}_{i,1}, \dots, \mathbf{b}_{i,c_i})^T \in \mathbb{R}^{3c_i}$ where $\mathbf{b}_{i,j} = (b^x_{i,j}, b^y_{i,j}, b^z_{i,j})^T$ are to model affine constraints on the trajectory, where $j \in \{1, \dots, c_i\}$, and c_i is the number of affine constraints on i^{th} k-space shot. The purpose of $\mathbf{A}_i$ is to select the portion of k-space shot where the constraints need to be active and vector $\underline{\mathbf{b}}_i$ defines the specified constraints, i.e. the locations in k-space to go through. Note that as compared to [Cha+16], here the set of affine constraints is generalized as its number $c_i > 1$ and the constraints themselves ($\mathbf{A}_i$, $\underline{\mathbf{b}}_i$) may vary across shots.

In [Cha+22d], a TE constraint was used, which ensures that each shot passes through the k-space center at TE. This is done to ensure that the same target contrast is measured across multiple shots crossing the center of k-space at the same time point, i.e. TE. The

index of k-space sample at TE being given by $1 \leq k_{\mathrm{TE}} = \left\lfloor \frac{\mathrm{TE}}{\Delta t} \right\rfloor \leq N_s$, this corresponds to a single constraint on every shot i, i.e. $c_i = 1$ with:

$$\mathbf{a}_{i,1}^{d}[k] = \begin{cases} 1 & k = k_{\mathrm{TE}}, \forall d \in \{x, y, z\}, \\ & \text{and } \forall i \in \{1, \ldots, N_c\} \\ 0 & \text{otherwise} \end{cases} \tag{5.2a}$$

$$\mathbf{b}_{i,i} = [0, 0, 0]^T \quad \forall i \tag{5.2b}$$

This way the constraint $\mathbf{A}_i \mathbf{k}_i = \underline{\mathbf{b}}_i$ ensures that at TE, the k-space samples $\mathbf{k}_i[k_{\mathrm{TE}}] = [0, 0, 0]^T$, $\forall i = 1 \ldots N_c$. More sophisticated linear constraints like gradient moment nulling can be modeled through $\mathbf{A}_i$ and $\underline{\mathbf{b}}_i$, by not limiting entries in $\mathbf{A}_i$ to be binary (see details in [Cha+16]).

5.2.2 3D SPARKLING

From [Boy+16; Cha+17; Laz+19; Cha+22d] we optimize the 3D trajectory $\mathbf{K} \in \Omega^N$ with $N = N_c \times N_s$ sampling points using the SPARKLING algorithm:

$$\widehat{\mathbf{K}} = \arg \min_{\mathbf{K} \in \mathcal{Q}_{\mathbf{A}, \underline{\mathbf{b}}^{N_c}}} F_N(\mathbf{K}; \Pi) = \left[F_N^{\mathrm{a}}(\mathbf{K}; \Pi) - F_N^{\mathrm{r}}(\mathbf{K}) \right] \tag{5.3}$$

where $F_N^{\mathrm{a}}(\mathbf{K}; \Pi)$ is the attraction term which ensures the sampling pattern $\mathbf{K}$ follows a prescribed TSD Π and $F_N^{\mathrm{r}}(\mathbf{K})$ the repulsion term to avoid clustering of samples. From [Cha+22d; Laz+19]:

$$F_N^{\mathrm{a}}(\mathbf{K}; \Pi) = \frac{1}{N} \sum_{n=1}^{N} \int_{\Omega} \|x - \mathbf{K}[n]\|_2 \, \Pi(x) dx \,, \tag{5.4a}$$

$$F_N^{\mathrm{r}}(\mathbf{K}) = \frac{1}{2N^2} \sum_{1 \leq n, n' \leq N} \|\mathbf{K}[n] - \mathbf{K}[n']\|_2 \,. \tag{5.4b}$$

The sampling pattern $\mathbf{K}$ is optimized using projected gradient descent algorithm, as described in [Cha+22d, Algorithm 1].

In practice, the optimization is performed through multi-resolution (see [Cha+22d, Sec.II-E]) which starts by spreading $N_{R^{\max}} = N/2^{R^{\max}}$ samples at the maximal $R^{\max} = 5$ decimation levels and iterates through a dyadic process, i.e. $N_{R^{\max}-R} = 2^R N_{R^{\max}}$ for $R = 1$ to 5 ($N_0 = N$). This is performed to ensure that the optimization is carried out with faster convergence when we coarsely optimize the k-space trajectory initially ($R = R^{\max}$). Then optimization is refined at finer resolutions as we approach convergence ($R = 1$).

5.3 Methods

In this section, we briefly describe two major extensions of SPARKLING that provide improved reconstructed image quality with reduced off-resonance artifacts.

5.3.1 MORE-SPARKLING

From [Fes10; Don20], the measured k-space samples $\mathbf{Y} = (\mathbf{y}_i)_{i=1}^{N_c}$ across the N_c shots are given by:

$$\mathbf{y}_i(t) = \int_{FOV} \mathbf{x}_{\boldsymbol{r}} \, e^{-(\alpha_{\boldsymbol{r}} + \imath \omega_{\boldsymbol{r}})t} \, e^{-2\imath\pi(\mathbf{k}_i(t) \cdot \boldsymbol{r})} \, d\boldsymbol{r} \tag{5.5}$$

with $\mathbf{x}_{\boldsymbol{r}}$ the transverse magnetization of the object, $\alpha_{\boldsymbol{r}}$ the T_2^* decay and $\omega_{\boldsymbol{r}}$ the off-resonance at voxel $\boldsymbol{r}$. Note that the temporal dependence of $\mathbf{Y}$ on $\alpha_{\boldsymbol{r}}$ and $\omega_{\boldsymbol{r}}$, was not considered in the original SPARKLING formulation. Due to this, the SPARKLING trajectories result in a sampling pattern where multiple samples present nearby in k-space are collected at

different times, thereby inducing artifacts. To observe this, we present temporal sampling plot for mid-plane of k-space along z-axis in Fig. 5.1(A), where we use rainbow coloring scheme to show the time at which the k-space sample is collected in each shot. We see that in region marked with blue arrow, for conventional SPARKLING trajectories, the k-space is not smooth temporally, resulting in trajectories that may be locally inconsistent leading to amplified ΔB_0 artifacts.

We mitigate the impact of B_0 inhomogeneities by adding temporal weights in the repulsion term $F^{\mathrm{r}}_N(\mathbf{K})$ to obtain $F^{\mathrm{r}}_{N,\tau}(\mathbf{K})$:

$$F^{\mathrm{r}}_{N,\tau}(\mathbf{K}) = \frac{1}{2N^2} \sum_{1\leq n,n'\leq N} e^{\left(\frac{|t_n - t_{n'}|\cdot\tau}{2^{R^{\max}-R}}\right)} \|\mathbf{K}[n] - \mathbf{K}[n']\|_2 \tag{5.6}$$

where k-space sample location $\mathbf{K}[n]$ is sampled at time $t_n = n\Delta t$ after the RF pulse during Observation time (T_{Obs}) and $\tau \geq 0$ is a scalar user-defined repulsion weighting parameter. The purpose of the weighting $\frac{1}{2^{R^{\max}-R}}$ is to shape the amount of temporal repulsion added as a function of the current decimation level, where we have a stronger temporal repulsion at initial stages of the algorithm. As we approach convergence and finer resolution levels (lower R), the amount of temporal weighting is significantly reduced to prevent the presence of unwanted k-space holes.

Notice that when $t_n \approx t_{n'}$, then $F^{\mathrm{r}}_{N,\tau} \approx F^{\mathrm{r}}_N$, and we are still solving the original problem in Eq. (5.3). In contrast, when $t_n \neq t_{n'}$ and $\tau \geq 0$, then $F^{\mathrm{r}}_{N,\tau} > F^{\mathrm{r}}_N$, thereby the sample points are facing a stronger repulsion, pushing them apart. This way, the k-space locations which are sampled at different time points are pushed apart, thereby ensuring a smooth temporal sampling of the k-space while also satisfying the TSD Π. These updated k-space sampling trajectories are presented in Fig. 5.1(B)-(F) and moving forward, we call these trajectories as MORE-SPARKLING . Note that when $\tau = 0$, $F^{\mathrm{r}}_{N,\tau} = F^{\mathrm{r}}_N$ and the resulting trajectories are the conventional SPARKLING .

With this, we generate MORE-SPARKLING trajectories, with TSD which takes the TE constraints into account (see [Laz+19, Sec. 2.3]) for varying τ values and present them in Fig. 5.1(A)-(F). We observe that strong weighting of this repulsion term with an increased τ results in k-space holes (marked by red arrows), which is detrimental to optimal reconstructed image quality. To prevent this, τ needs to be grid-searched appropriately to enforce temporally smooth k-space sampling, while avoiding undesirable k-space holes. The effect of varying τ is presented in Fig. 5.1. In our studies (Sec. 5.3.1), we observed that $\tau = 1.0$ (Fig. 5.1D) resulted in maximal signal recovery with minimal impact of the k-space holes on final reconstructed image quality.

5.3.2 GoLF-SPARKLING

Another concern of traditional SPARKLING trajectories is the presence of TE constraints on the k-space sampling trajectories. These constraints are specifically added to the SPARKLING formulation to ensure that the center of k-space or the low frequencies are sampled at the same time after the RF pulse (i.e. at TE) as described in Sec. 7.2.7.

Consequently, resulting k-space trajectories sample the central frequencies at a rate higher than Nyquist rate, resulting in multiple samples collected within a Cartesian k-space voxel. Although such oversampling is fully justified by CS theory and VDS, in practice it leads to accumulation of ΔB_0 artifacts as multiple samples are acquired in the same Cartesian k-space voxel through an integration over different k-space trajectory paths (see Eq.(5.5)). Instead, these samples could be dispatched in other portions of k-space (e.g. in higher frequencies) to increase its coverage at no additional cost and provide more details in the images.

This oversampling of the center can be tackled partially by updating the TSD by taking the TE point constraint into account (see [Laz+19, Sec. 2.3]) and reducing the density around the center of k-space to limit the number of samples per Cartesian k-space voxel. However, we don't fully mitigate this problem as we still end up with N_c k-space sample points at the center of k-space $[0, 0, 0]^T$ causing nearby regions to be still sampled densely. This problem cannot be addressed plainly with removal of TE point constraint and using temporal weights

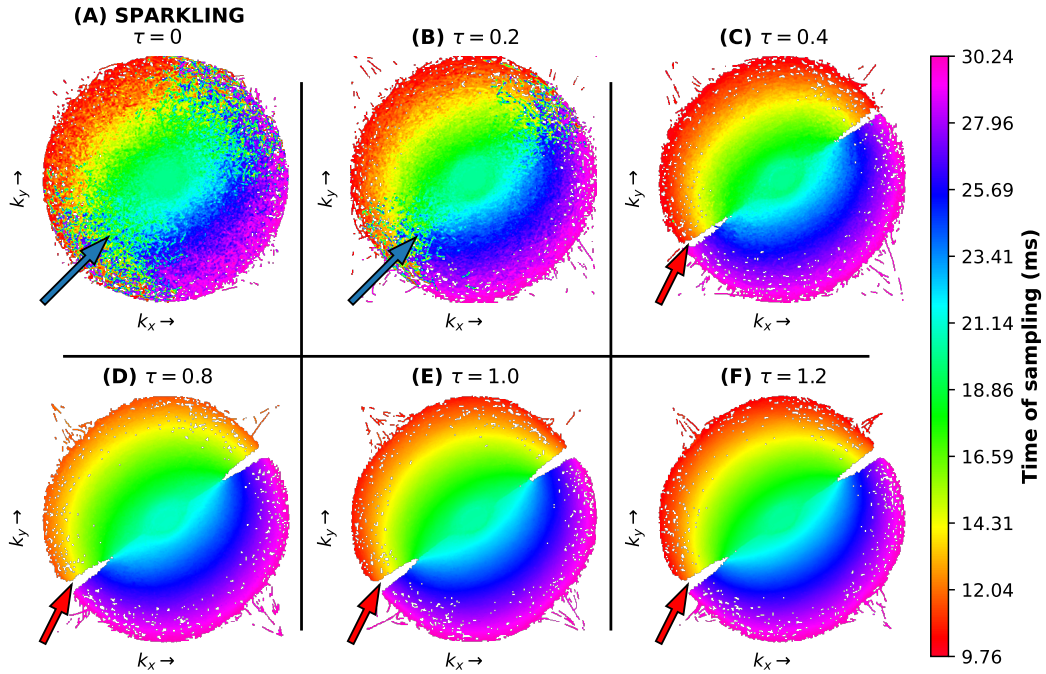


Figure 5.1 – **Comparison of different MORE-SPARKLING trajectories with varying temporal weights (τ):**
Comparison of different SPARKLING trajectories generated with $N_c = 3639$ (AF=20), $N_s = 2048$ (A) without temporal weights ($\tau = 0$) i.e. original SPARKLING , and with τ varying from 0.2 to 1.2 as shown from (B) to (F), respectively. A rainbow coloring scheme overlays the sampling trajectories to encode the time over k-space samples.

(from Sec. 5.3.1) to achieve temporally smooth sampling within Nyquist criteria, as then the resulting sampling pattern would not sample the center of k-space at TE to obtain the desired target contrast.

In order to efficiently mitigate such issues of oversampling, note that an optimal way to sample a region of k-space at Nyquist with minimum redundancy is through Cartesian sampling. We use this fact and add affine constraints into our projection set $\mathcal{Q}^{N_c}_{\mathbf{A},\underline{\mathbf{b}}}$ such that we carry out Cartesian sampling at center of k-space. Each k-space sample shot $\mathbf{k}_i, \forall i \in \{1, \ldots, N_c\}$, is constrained to pass through the lower frequencies in the form of a Cartesian line as shown in Fig. 5.2. This constraint is enforced by crafting individual $\mathbf{A}_i$ and $\mathbf{b}_i$ in a specific manner for each shot i.

In practice, we cover a sphere $\mathcal{S}$ in the center of k-space defined in Ω with Cartesian sampling as shown in Fig. 5.2(A) where we present GoLF-SPARKLING trajectories for $N_c = 256$ and $N_x = N_y = N_z = \tilde{N} = 64$. The lower resolution and number of shots is particularly chosen for better visualization. The entire k-space trajectory can be split into a non-Cartesian part (blue, Fig 5.2(B)) and a Cartesian part (green, Fig 5.2(C)). Further, for mathematical simplicity, we assume the image matrix sizes to be equal, i.e. $\tilde{N} = N_x = N_y = N_z$, while extensions to non-isotropic matrix sizes can be carried out as the k-space sampling space is normalized in $\Omega \in [-1, 1]^3$. Particularly, we cover $\mathcal{S}$ with straight line readouts along x as shown in Fig. 5.2(E), and cover a circle of radius $r_{\mathcal{S}}$ in k_y and k_z encoding directions as seen in Fig. 5.2(D). Thus, the pixel size of the Cartesian grid in k-space is given by $2/\tilde{N}$. As each k-space shot forms a Cartesian line, from Fig. 5.2(F) we obtain the area in the central slice with Cartesian sampled circle as:

$$\text{Area} = \pi r_{\mathcal{S}}^2 \simeq \left(\frac{2}{\tilde{N}}\right)^2 N_c \,.$$

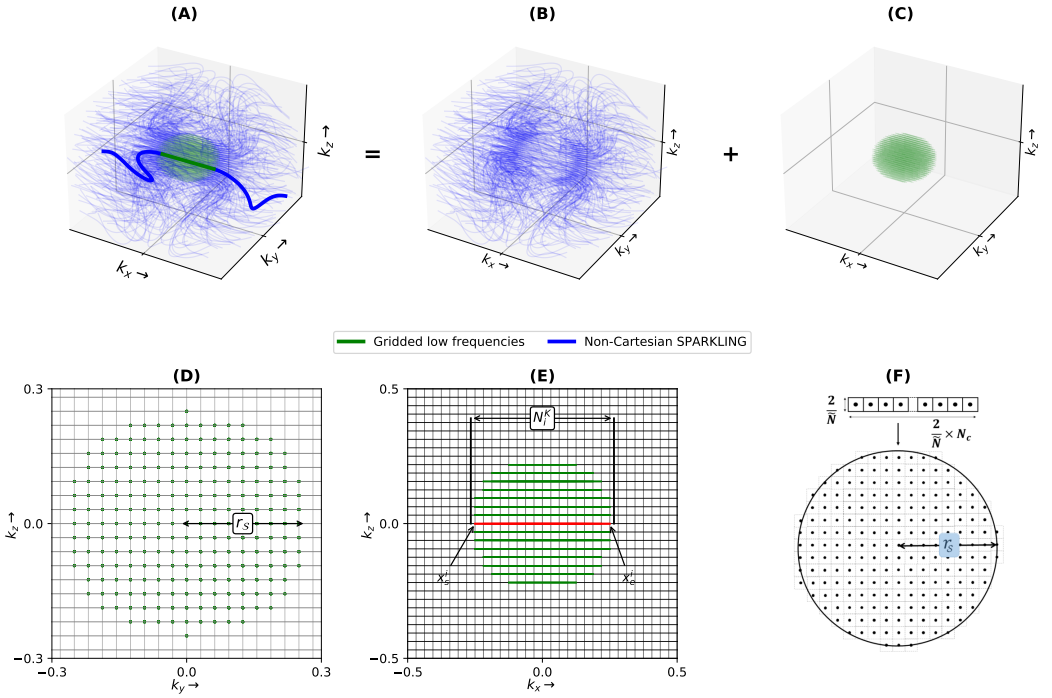


Figure 5.2 – **GoLF-SPARKLING trajectories for $N_c = 256$, with $\tilde{N} = 64$ (for clearer visualization):**
(A) GoLF-SPARKLING trajectory which is composed of **(B)** non-Cartesian SPARKLING portion in blue and **(C)** gridded low frequencies in green. Slice profile of the Cartesian portion of the k-space trajectory is presented along **(D)** $k_x = 0$ plane and **(E)** $k_y = 0$ plane. **(F)** The ordering of N_c k-space shots in sphere $\mathcal{S}$ at Nyquist criteria.

With this, we obtain $r_{\mathcal{S}} = \frac{2}{\tilde{N}} \left\lfloor \sqrt{\frac{N_c}{\pi}} \right\rfloor$. Based on Nyquist criteria, we need to sample at least every $\Delta x = \frac{2}{\tilde{N}}$ to have non-aliased reconstructed image. However, as the k-space data is sampled at every dwell time $\delta t < \Delta t$, in practice, the scanner can play k-space trajectory to take at least $\frac{\Delta x \Delta t}{\delta t}$ samples at every Δt. Generalizing this, we introduce k-space velocity at center of k-space as a dimensionless parameter $v \geq 0$, which is the number of Nyquist voxel steps Δx taken in readout direction per gradient raster time Δt. Particularly, k-space trajectory takes steps of size $v\Delta x < \frac{\Delta x \Delta t}{\delta t}$ at every Δt. The purpose of v is to efficiently utilize the gradient hardware in traversing the center of k-space at maximum feasible speed while maintaining Nyquist criteria after sampling at ADC at every δt.
Say the Cartesian portion of i^{th} k-space shot $\mathbf{k}_i$, starts as a k-space line at x_i^s and ends at x_i^e, then we obtain the number of samples at center of k-space for this shot as follows:

$$N_i^K = \left\lceil \frac{x_i^e - x_i^s}{v\Delta x} \right\rceil . \tag{5.7}$$

The k-space locations of these Nyquist points are given by:

$$x_i^\ell = x_i^s + \frac{\ell\,(x_i^e - x_i^s)}{N_i^K}, \quad \forall \ell \in \{1, \ldots, N_i^K\} . \tag{5.8}$$

This results in $c_i = N_i^K$ constraints which are applied to i^{th} k-space shot $\mathbf{k}_i$ at N_i^K indices between $k_i^L = \left\lfloor \frac{N_c - N_i^K}{2} \right\rfloor$ and $k_i^H = \left\lfloor \frac{N_c + N_i^K}{2} \right\rfloor$, where L and H stand for lower and higher

indices, respectively. With this, Eq. (5.2) updates as follows:

$$\mathbf{a}_{i,j}^{d}[k] = \begin{cases} 1 & \text{for } k = k_i^L + j \quad j \in \{1,\dots,c_i\} \\ & i \in \{1,\dots,N_c\} \quad d \in \{x,y,z\} \\ 0 & \text{otherwise} \end{cases} \tag{5.9a}$$

$$b_{i,\ell} = x_i^{\ell} \quad \forall \ell \in \{1,\dots,N_i^K\}\,. \tag{5.9b}$$

Then, the affine constraints in matrix formulation in Eq. (5.2) read: $\mathbf{k}_i[k_i^L + \ell] = x_i^{\ell}, \forall \ell \in \{1,\dots,N_i^K\}$ match the affine constraints in $\mathcal{Q}_{\mathbf{A},\underline{\mathbf{b}}}^{N_c}$. We call this update to SPARKLING as Gridding of Low Frequencies or GoLF.

5.3.3 TSD characterization for GoLF

In this section, we derive the mathematical expression for the TSD in k-space for the particular case of GoLF-SPARKLING , such that we enforce Nyquist sampling at the center of k-space. It is important to note that as we add constraints on k-space trajectories in the center of k-space, the corresponding TSD is also affected. For simplicity, we derive the TSD for the specific case of $v = 1$ i.e. we sample 1 Nyquist voxel every Δt. However, in Sec. B.1, we derive the TSD for the general case of v in detail.

Following [Cha+22d], we parameterize the TSD as radially isotropic which decays at an inverse polynomial rate D and reaches a constant plateau in center of k-space up to a cutoff frequency C. For the case of GoLF, we set $C = r_{\mathcal{S}}$ and parameterize the TSD in k-space as follows:

$$\Pi_{r_{\mathcal{S}},D}(x) = \begin{cases} \kappa & |x| < r_{\mathcal{S}} \\ \kappa \left(\frac{r_{\mathcal{S}}}{|x|}\right)^D & |x| > r_{\mathcal{S}} \end{cases} \tag{5.10}$$

where κ, the normalizing constant is the density of the plateau, which for GoLF trajectories must match the density for Nyquist criteria.

For a total number of gradient raster sampling points $N = N_c \times N_s$, the number of samples within the center of k-space ($|x| < r_{\mathcal{S}}$) is given by $N_{|x|<r_{\mathcal{S}}}$:

$$N_{|x|<r_{\mathcal{S}}} = N\kappa\frac{4}{3}\pi r_{\mathcal{S}}^3\,. \tag{5.11}$$

With Cartesian sampling in the center of k-space, we sample once per Nyquist voxel of side length Δx. Thus, the number of Cartesian Nyquist sample points N_{Nyq} in the center of k-space is given by:

$$N_{\text{Nyq}} = \frac{\frac{4}{3}\pi r_{\mathcal{S}}^3}{\Delta x^3}\,. \tag{5.12}$$

Then the Nyquist sampling criteria are enforced for the center of k-space by setting $N_{|x|<r_{\mathcal{S}}} = N_{\text{Nyq}}$ to get:

$$\kappa = \frac{1}{N\Delta x^3}\,. \tag{5.13}$$

Finally, as $\Pi_{r_{\mathcal{S}},D}(x)$ is a distribution, we need to ensure that it is normalized (i.e. sum to 1) giving us (see B.1 for details):

$$\kappa 4\pi\left(\frac{r_{\mathcal{S}}^3}{3} + \frac{(r_{\mathcal{S}})^D - r_{\mathcal{S}}^3}{3-D}\right) = 1\,. \tag{5.14}$$

With this, the decay D can be obtained by solving Eq. (5.14) iteratively, and we present it and the corresponding $r_{\mathcal{S}}$ at the top of Fig. 5.3, as a function of AF specified by AF$= \frac{N_y \times N_z}{N_c}$. Additionally in Fig. 5.3, we see that D increases and $r_{\mathcal{S}}$ (i.e. C) decreases as AF increases, hence higher frequencies are more sparsely sampled resulting in a more peaky TSD for higher

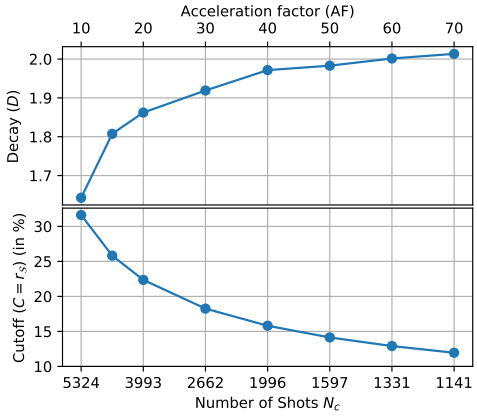


Figure 5.3 – Characterization of cutoff C and decay D for GoLF-SPARKLING trajectories with $v = 1$ and varying AF (specified at top) and its corresponding number of trajectories (specified at bottom).

AF values. However, in the asymptotic regime, D is nearly 2.2, while even at high AF=70, prescribed by $N_c = 1141$, we still have $r_{\mathcal{S}} = 12\%$ of center of k-space sampled on a grid.

In the case of NC sampling at Nyquist rate in a region, we need to use Poisson disk sampling. Particularly in 3D, this is equivalent to filling up the k-space with a sphere of diameter Δx (rather than a Nyquist voxel of side length Δx). Hence, as already known [BKZ04b, Chap. 13] with NC sampling we need more samples to satisfy the Nyquist criteria as compared to Cartesian sampling. The ratio of number of Nyquist samples with 3D NC imaging (N^{nc}) to the case of 3D Cartesian imaging (N^c) is given by the inverse of ratio of the corresponding volumes:

$$\frac{N^{nc}}{N^c} = \frac{\Delta x^3}{\frac{4}{3}\pi\left(\frac{\Delta x}{2}\right)^3} = \frac{6}{\pi}\,. \tag{5.15}$$

Note that at the edges where we switch from Cartesian sampling to NC sampling, we need a larger target density to ensure we still satisfy Nyquist criteria. We take this into account in our detailed generic formulation in Sec. B.1.

Finally, we can apply 3D Inverse fast Fourier transform (IFFT) on the k-space data within the gridded sampling region (see Fig. 5.2(C)) to obain an artifact-free low resolution version of the 3D MR images. Such volumes can be computed for all coils to estimate the coil sensitivity maps through self-calibration [El +18b] without any additional scan, which can be instrumental in the image reconstruction process.

5.3.4 MORE + GoLF SPARKLING

In the above sections, we introduced 2 novel important features to the vanilla SPARKLING trajectories, which help in reducing artifacts and improving the reconstructed image quality. Observe that MORE feature involves only a change to the repulsion term $F^r_{N,\tau}$, GoLF involves a change to the constraint set $\mathcal{Q}^{N_c}_{\mathbf{A},\underline{\mathbf{b}}}$ and a corresponding change to the TSD (Π) and thereby the attraction term $F^a_{N,\Pi}$. Hence, MORE and GoLF features are totally independent and can be combined to form MORE+GoLF-SPARKLING trajectories with temporally smooth k-space sampling pattern and gridded sampling in center of k-space.

5.3.5 MRI acquisition parameters

We carried out *in silico* and *in vivo* measurements to validate and benchmark the improvements of the new trajectories MORE and GoLF. In all the experiments, the target resolution is 0.6mm^3 isotropic, with $N_x = N_y = 384$ ($\text{FOV}_\text{x} = \text{FOV}_\text{y} = 23$cm) and $N_z = 208$ ($\text{FOV}_\text{z} = 12.48$cm). We use an AF of 20 ($N_c = 3969$), except in the case of variable AF study. All the prospective scans were carried out on a clinical 3T MR system (Magnetom

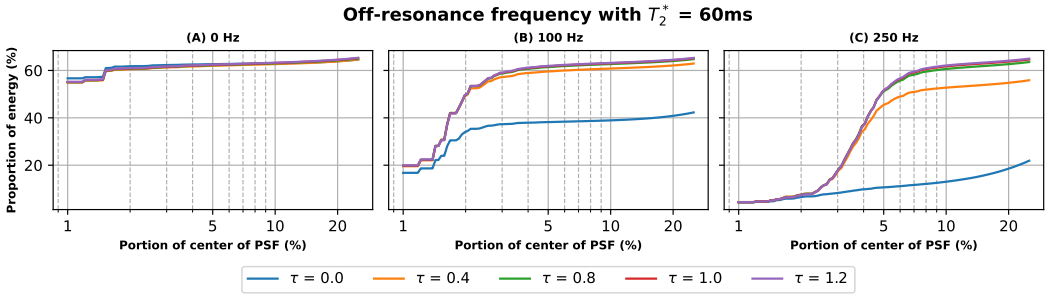


Figure 5.4 – **Proportion of total energy around the center of PSF for varying levels of ΔB_0 .**
The accumulated proportion of energy for MORE-SPARKLING trajectories with varying τ from 0 to 1.2. The energy of the 3D PSF within a radial shell with radii at different distances from center shown on x-axis on a log scale. The PSF is obtained with a constant T_2^*= 60ms and off resonance frequencies of **(A)** 0 Hz, **(B)** 100 Hz and **(C)** 250 Hz.

Prisma^{FIT}, Siemens Healthcare, Erlangen, Germany) where the k-space data was acquired using a Siemens 64 channel Head/Neck coil. T_2^*-w images were obtained with flip angle of 15° and slab selective excitation with TE of 20ms and TR of 37ms. The k-space data was acquired with T_{Obs}= 20.48ms ($N_s = 2048$, hence $\Delta t = 10$ µs) and data was sampled by the ADC at $\delta t = 2$ µs. Additionally, a ΔB_0 map was acquired with a 2D gradient echo sequence in the same FOV at 2mm isotropic resolution with $\text{TE}_1 = 4.92$ms and $\text{TE}_2 = 7.38$ms. These TEs enable the coverage of ΔB_0 inhomogenieties in range $[-203, 203]$Hz, resulting in 1 phase wrap present in all references which was unwrapped using [Her+02]. For *in silico* scans, we used the NIST/ISMRM MRI system phantom [21] for calibrations and testing as this phantom can be used for assessing geometry distortions, image uniformity and resolution. Our *in vivo* scans was done on one volunteer with approvals from local and national ethical committees for the protocol (CPP 100048), and after a written consent was obtained from the volunteer.

5.3.6 MR image reconstruction

All the reconstructions for the data from SPARKLING trajectories were carried out off-line using self calibrating MR reconstruction [El +18b] using pysap-mri[1], a plugin for PySAP[2] [Far+20a]. We used the synthesis formulation of self-calibrated CS reconstruction with ℓ_1-norm regularization in symlet-8 wavelet domain to promote sparsity. Optimization of this cost function was performed using Fast Iterative Shrinkage-Thresholding Algorithm (FISTA). To reduce computational complexity, k-space data was coil compressed to 10 channels using the principal component analysis-based method proposed by [Bue+07]. For Cartesian references, the corresponding magnitude and phase DICOMs were directly obtained from the scanner.

5.4 Results

In this section, we present the results for different experiments carried out for optimizing parameters in the new SPARKLING trajectories using MORE and GoLF features. Later, we proceed to present the overall improvement observed when combining MORE and GoLF features under optimal settings. All the proposed experiments were performed in a prospective validation setting i.e. in truly accelerated imaging scenarios.

[1]https://github.com/CEA-COSMIC/pysap-mri
[2]https://github.com/CEA-COSMIC/pysap

5.4.1 MORE: Optimizing τ

In this experiment we optimized the parameter τ for MORE-SPARKLING through two studies: (i) Analysis of the PSF and (ii) through prospective study.

PSF

We computed the PSF of the MORE-SPARKLING trajectories for different values of τ and show the results in Fig. 5.4. The details on how to compute the PSF are given in Sec. B.2. For optimal visualization of 3D PSF, we present accumulated energies within concentric radial shells with different radii as a percentage of total energy. These plots help us understand how peaky the PSF is and having a higher percentage of energy within a smaller central portion of k-space leads to minimal blurring and reduced artifacts in the reconstructed image. Varying levels of ΔB_0 were superimposed to the k-space data to understand the behavior of trajectories in different settings. For T_2^*-w imaging at 3T, the PSF was obtained with T_2^*= 60 ms (mean T_2^* value for gray and white matter [Pet+07])and the off resonance frequencies were chosen at 3 levels: (A) Ideal case with no ΔB_0 (0 Hz), (B) medium level of ΔB_0 (100 Hz) and (C) maximum level of ΔB_0 (250 Hz). We also present similar results for 1.5T and 7T in Fig. B-1 in Sec. B.2.

From Fig. 5.4 we see that for $\tau = 0$ (SPARKLING trajectory), the PSF strongly degrades when ΔB_0 increases, with only 20% of energy in the 20% of the central region. In contrast, we observe that with addition of temporal weighting ($\tau > 0$), this PSF is improved strongly with nearly 60% of energy within 4% (respectively, 8%) of the central region around the peak at 100 Hz (resp. at 250Hz). Further, we see that the improvement of the PSF with increase in τ beyond $\tau = 0.8$ is very limited and incremental. With this study, we conclude that $\tau \geq 0.8$ is the optimal setting for MORE SPARKLING the case of T_2^* imaging at 3T.

Prospective study

As described in Sec. 5.3.1, the amount of temporal repulsion or τ needs to be grid searched to obtain an optimal value which maximally reduces the impact of ΔB_0 without causing artifacts induced by k-space holes in sampling patterns. For this, we obtained MORE-SPARKLING trajectories with varying τ from 0 to 1.2, and a prospective study was run first on the NIST phantom (Fig. 5.5(1)) and then *in vivo* on a healthy volunteer (Fig. 5.5(2)).

For phantom, volumes were first collected in a standard acquisition setup with low artifacts (Fig. 5.5(1a)), to observe any loss in image quality due to k-space holes emerging from large τ. To understand the extent to which MORE-SPARKLING trajectories can mitigate ΔB_0 artifacts, strong B_0 inhomogeneities were added by degrading the machine B_0 shimming with spherical harmonics (Fig. 5.5(1b)). The prospective study on phantom (Fig. 5.5) shows significant improvements over B_0 inhomogeneities where an increase in τ results in minimized blurring and signal recovery in regions marked with green arrow (see Fig. 5.5(1b)F).

Further, we tested the same trajectories in the *in vivo* environment and present results along mid axial plane (Fig. 5.5(2a)) and mid-sagittal plane (Fig. 5.5(2b)). Maximal signal losses are seen for SPARKLING trajectories shown with red arrow as shown in Fig. 5.5(C). Additionally, we clearly see a gradual signal recovery when increasing τ in the same marked regions. Finally, we observe that $\tau = 1.0$ is optimal in terms of maximal signal recovery with minimal k-space holes (from Fig. 5.5(F)).

In both degraded acquisition settings, Fig. 5.5 demonstrates that accelerated Cartesian imaging (GRAPPA-4) is more robust to ΔB_0 inhomogeneities than MORE-SPARKLING due to increased temporal smoothness (i.e. imposed by a one-dimensional readout) in Cartesian sampling. However, these Cartesian scans are 6.3-fold longer.

5.4.2 GoLF: Varying trajectory velocity at the center of k-space

Based on Eq. (5.7), the velocity of the Cartesian portion of the k-space trajectory at the center of k-space can be controlled using v. Under the maximum gradient strength of

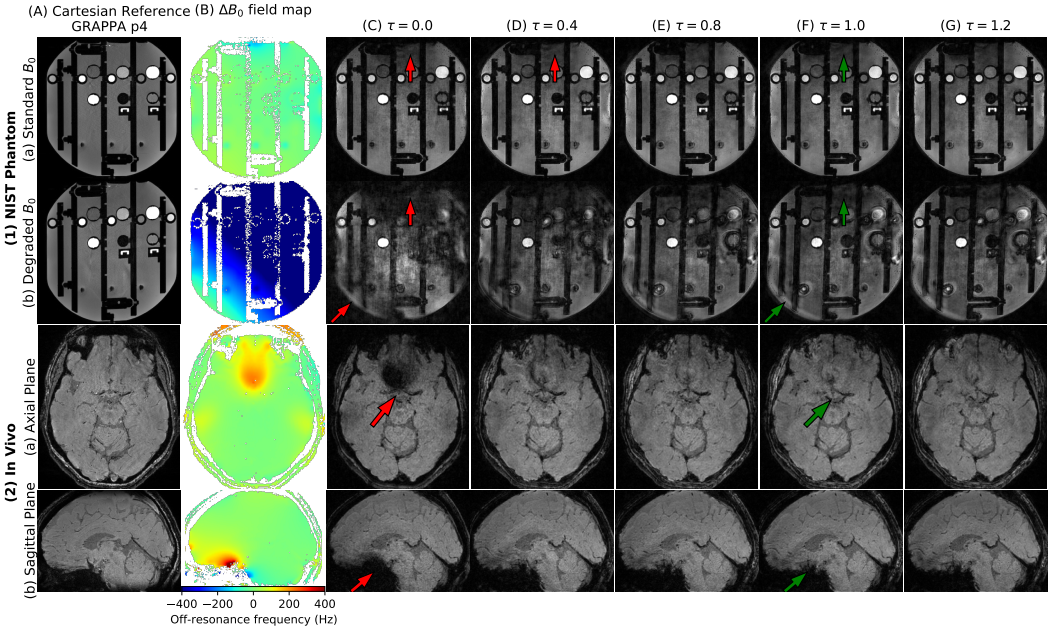


Figure 5.5 – **Prospective results with varying temporal weights (τ) on phantom and *in vivo* (brain imaging):**
Different trajectories are compared on **(1)** NIST phantom with **(a)** standard B_0 and **(b)** added B_0 inhomogeneities by degrading the shim and **(2)** *in vivo* volunteer along **(a)** axial and **(b)** sagittal planes. SPARKLING ($\tau = 0$) and MORE-SPARKLING acquisitions are carried out at AF=20 (scan time = 2min30sec) and for varying values of τ as shown in the figure **(C)-(G)**. Further we present **(A)** GRAPPA 4-fold (p4) acquisition as Cartesian reference (scan time = 15min30sec) and **(B)** the corresponding ΔB_0 field map. The regions with maximum degradation due to ΔB_0 is marked with red arrows in **(C)** and the regions with improvement in signal recovery and image quality are marked with green arrows in **(F)**.

40mT/m and the acquisition parameters as described in Sec. 5.3.5, a maximum of 3.92 Nyquist steps can be taken by the MR scanner in a given Δt (see Sec. B.3). High velocities at center of k-space is useful to result in longer k-space trajectory, which leads to larger k-space coverage. However, in the prospective setting, traversing the center of k-space at higher velocity may lower the SNR as the MR signal in k-space is accumulated over a shorter time interval. Consequently, there is a need to find an optimal value of v that reaches the best trade-off between an improved k-space coverage and a limited SNR loss to avoid degrading too much image quality.

Prospective scans were done on the NIST phantom for GoLF-SPARKLING trajectories with $v = 0.6, 1, 2$ and 3 (Fig. 5.6). The NIST phantom can help optimally tune v as it embodies resolution insets, which can be used to quantify image quality (see [Cha+22d, Sec IV.D2]). Overall, in the axial plane, Fig. 5.6 shows an increase in artifact level appearing first at $v = 2$ and becoming more prominent for $v = 3$, as marked with red arrow in the bottom row. Further, when zooming into the resolution insets, for $v > 1$, we see significant losses in the details. This careful analysis indicates that $v \leq 1$ is better for T_2^*-w imaging at 3T. Within this range, for the sake of broader k-space coverage we chose $v = 1$, i.e. the highest possible velocity to spend minimum time in the center of k-space and collect more samples in the high frequency region, resulting in sharper edges (green arrow in the bottom row).

5.4.3 Joint MORE and GoLF SPARKLING

In this section, we demonstrate the overall gain obtained when combining MORE and GoLF to minimize ΔB_0 artifacts and improve image quality through gridded sampling at center of

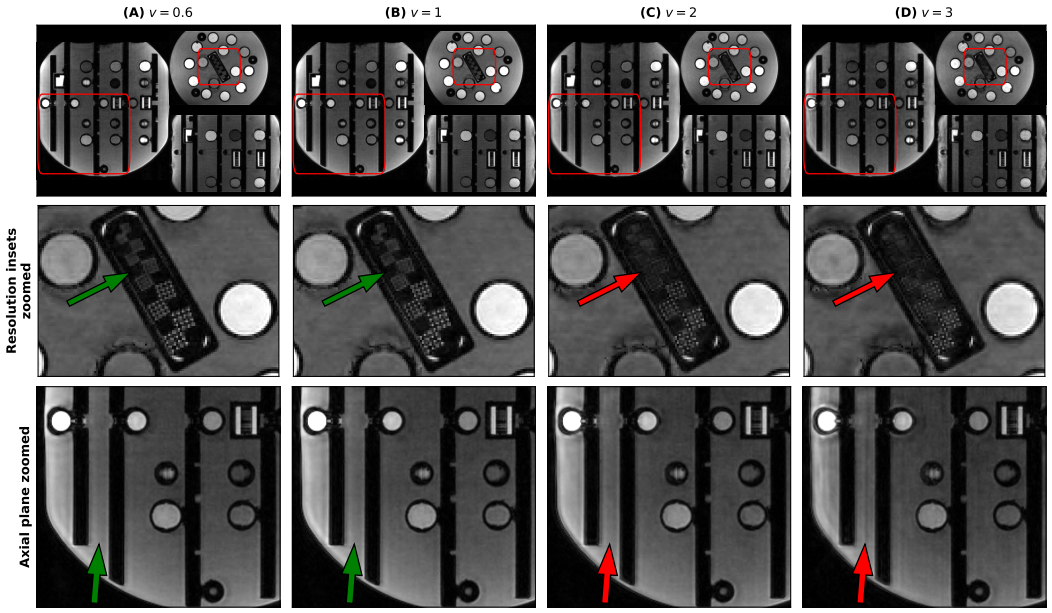


Figure 5.6 – Prospective results for GoLF-SPARKLING with varying trajectory velocity at the center of k-space on NIST phantom with k-space velocity as **(A)** $v = 0.6$, **(B)** $v = 1$, **(C)** $v = 2$ and **(D)** $v = 3$. We show slices from each orientation (top row), the zoomed in region into the resolution insets (mid-row) and zoomed in region in axial plane (bottom row).

k-space. For doing so, in Fig. 5.7 we present the *in vivo* results at $AF = 20$ with SPARKLING , MORE-SPARKLING , and finally their combination GoLF+MORE-SPARKLING .

As predicted, MORE-SPARKLING appears more robust than classic SPARKLING to ΔB_0 inhomogeneities and present minimized artifacts. With GoLF+MORE-SPARKLING we observe overall a clear gain in image quality with significant reduction in noise levels and closer contrast to the Cartesian reference, paving the way towards improved clinical use of the SPARKLING trajectories in the near future.

5.4.4 Varying acceleration factors (AF)

Prospective k-space data was also acquired *in vivo* from the same healthy volunteer at varying AF values from 15 to 50 for both MORE-SPARKLING and GoLF+MORE-SPARKLING trajectories. The aim of this experiment was to assess how image quality evolves as a function of scan time. The results are presented in Fig. 5.8. Overall, we observe that GoLF feature is crucial and provides less noisy and more detailed images as compared to the sole MORE-SPARKLING trajectories. The image quality is preserved up to AF=20 for GoLF+MORE-SPARKLING trajectories while we observe degradation at AF=20 for MORE-SPARKLING . Additionally, a direct diagonal comparison can be drawn between the two approaches: We observe that image quality at AF=20, AF=30, AF=40 and AF=50 for GoLF+MORE-SPARKLING trajectories is comparable to that of MORE-SPARKLING trajectories at AF=15, AF=20, AF=30 and AF=40, respectively.

5.5 Discussion and Conclusions

Arbitrary readout directions in non-Cartesian MRI leads to improved k-space coverage, but may cause increased off-resonance artifacts due to accumulation of ΔB_0 from different arbitrary trajectory paths. This major issue was observed in SPARKLING trajectories, particularly in the case of T_2^*-w imaging due to larger T_{Obs} (i.e. readouts) and TEs. In this work, we introduced two important features for SPARKLING trajectories to result in improved reconstructed images: MORE and GoLF.

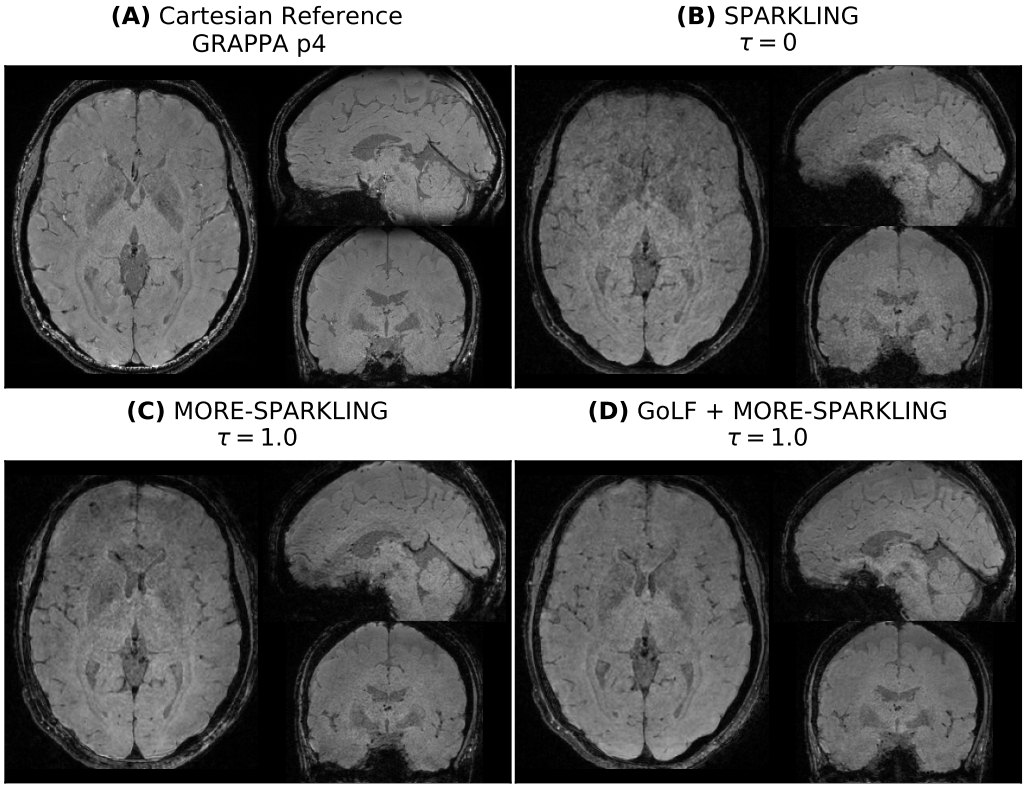


Figure 5.7 – Comparison of prospective results for on healthy volunteer at AF=20 (scan time = 2min30sec) with **(B)** conventional SPARKLING , **(C)** MORE-SPARKLING and **(D)** GoLF + MORE-SPARKLING trajectories. We have also presented **(A)** a Cartesian reference obtained with 4-fold GRAPPA acceleration (scan time = 15min30sec).

Through MORE, we introduced a temporal weighting in the traditional SPARKLING optimization problem which takes the temporal nature of the sampled data into account. The optimized trajectories resulted in temporally smoother k-space sampling, which led to reduced off-resonance artifacts with limited blurring and strong signal recovery in regions with large ΔB_0 . This temporal weighting can also be added to other k-space sampling pattern optimization problems, making this feature more generic and applicable widely to non-Cartesian MRI. For instance, these temporally smooth sampling patterns are extremely beneficial for X-Nuclei imaging where the values of T_2 relaxation times are small. MORE-SPARKLING trajectories sample a given region of k-space at the same time during scan leading to a stable signal level and overall an improved image quality. Preliminary data shows that center-out version of MORE-SPARKLING with shorter T_{obs} and utra-short TE were applied to Sodium Imaging in [Bap+22b] and were shown to outperform the state-of-the-art twisted projection imaging (TPI) [Boa+97] with clearer visibility in brain structure. In the latter context, further comparison with improved FLORET trajectories [Pip+11b; RAP17] would be insightful.

Cartesian sampling can sample a given region of k-space with minimum redundancy when sticking to the Nyquist criteria. Also, such sampling results in structured off-resonance artifacts, which can be corrected with simple post-processing. In GoLF, we incorporated these features of Cartesian sampling into the SPARKLING framework through more general affine constraints and adaptation of the TSD to match these Nyquist criteria constraints. The result was a significant decrease in artifacts and overall an improved image reconstruction quality.

Through this work, we introduced a novel compound sampling approach to measure the k-space with trajectories having both Cartesian and non-Cartesian parts to extract the best

of both worlds. With the GoLF feature, through Cartesian sampling at center of k-space, we can quickly obtain sensitivity maps in all our scans through simple IFFT of this central k-space data. As an extension to this, we can now incorporate parallel imaging methods like GRAPPA [Gri+02], SENSE [Pru+99] and CAIPIRINHA [Bre+05] to further increase the AF or increase the percentage of center of k-space sampled with Cartesian sampling, resulting in further improved image clarity.

Another extension to GoLF involves having the same k-space trajectory passing through the center of k-space as Cartesian line multiple times, resulting in a larger portion of center of k-space sampled with Cartesian sampling. This coupled with a high trajectory velocity (like $v = 3$), results in trajectories closer to echo planar imaging (EPI). Such trajectories can be helpful in achieving extremely high AF, which is crucial for imaging modalities like functional MRI. These extensions and specific applications to different modalities will be addressed in future works.

Concluding, we applied the above two features to SPARKLING framework, resulting in improved reconstructed image quality with reduced off resonance artifacts and noise level as well as clearer visibility in the structures. Using both features in conjunction allows us to accelerate the scans at unprecedented speeds, enabling to reach higher AF with significantly reduced degradation in image quality. As of now, we can speed up scans by 6.3 times compared to GRAPPA-4, leading to a 600 µm isotropic resolution scan in 3D T_2^*-w imaging possible in just 2.5 minutes at 3T with negligible degradation in image quality.

* * *
* *
*

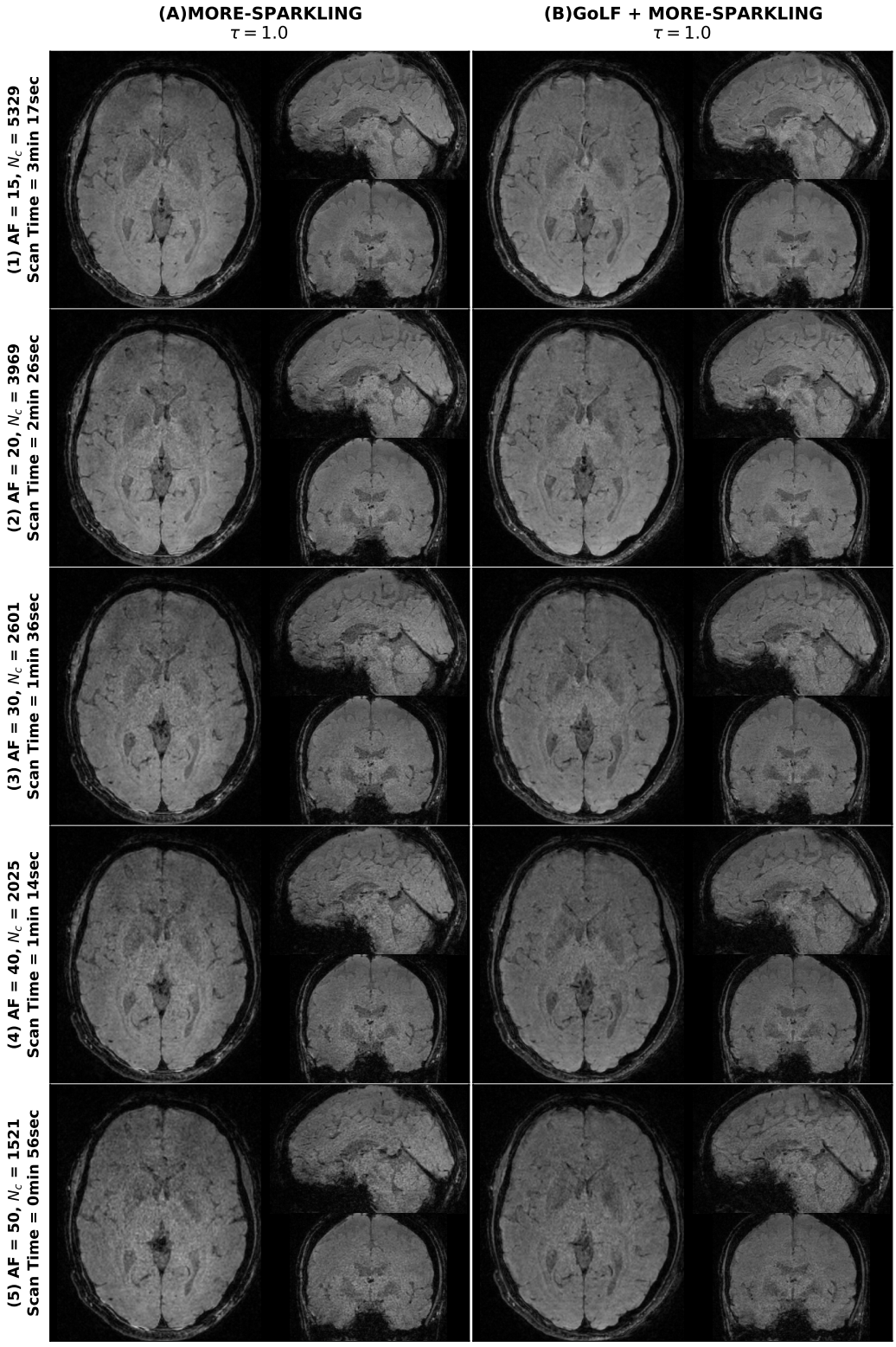


Figure 5.8 – **Varying AF study on prospective *in vivo*:**
Comparison of prospective results at varying AF on a healthy volunteer with (A) MORE-SPARKLING and (B) GoLF + MORE-SPARKLING trajectories with $\tau = 1$ at (1) AF=15, (2) AF=20, (3) AF=30, (4) AF=40 and (5) AF=50. The corresponding number of shots and scan times are also reported in the y-axis.

Chapter 6

Learning sampling density for 2D SPARKLING

This chapter covers content that has been published in a peer-reviewed Conference:

- **Chaithya, G R**, Z. Ramzi and P. Ciuciu. "Learning the sampling density in 2D SPARKLING MRI acquisition for optimized image reconstruction". In: *29th European Signal Processing Conference (EUSIPCO)*. Dublin, Ireland, Sept. 2021, pp. 960–964

The SPARKLING algorithm was originally developed for accelerated 2D magnetic resonance imaging (MRI) in the compressed sensing (CS) context. It yields non-Cartesian sampling trajectories that jointly fulfill a target sampling density while each individual trajectory complies with MR hardware constraints. However, the two main limitations of SPARKLING are first that the optimal target sampling density is unknown and thus a user-defined parameter and second that this sampling pattern generation remains disconnected from MR image reconstruction thus from the optimization of image quality. Recently, data-driven learning schemes such as LOUPE have been proposed to learn a discrete sampling pattern, by jointly optimizing the whole pipeline from data acquisition to image reconstruction. In this work, we merge these methods with a state-of-the-art deep neural network for image reconstruction, called XPDNet, to learn the optimal target sampling density. Next, this density is used as input parameter to SPARKLING to obtain 20x accelerated non-Cartesian

trajectories. These trajectories are tested on retrospective compressed sensing (CS) studies and show superior performance in terms of image quality with both deep learning (DL) and conventional CS reconstruction schemes.

6.1 Introduction

Compressed sensing (CS) in MRI [Lus+07] has led to a large reduction in scan time while maintaining a reasonable reconstructed MR image quality. Practically, CS is implemented by undersampling pseudo-randomly the k-space according to a variable sampling density [Puy+11; Cha+13; Cha+14; Adc+17; Boy+19]. The sampling pattern may be composed of multiple individual Cartesian lines (Cartesian Sampling), in which case variable density sampling (VDS) is implemented only along the phase encoding dimension. To go to higher reduction in scan times, non-Cartesian sampling is really helpful as it permits the implementation of 2D VDS with the help of non-Cartesian trajectories, such as radial spokes [Lau73] and spiral interleaves [Ahn+86a]). Although radial and spiral sampling are widespread, they are not really optimal as radial spokes don't cover the k-space perfectly and spiral interleaves do not exactly match a prescribed sampling density. Hence, severe artifacts impede image quality during CS reconstruction.

The Spreading Projection Algorithm for Rapid K-space samplING, or SPARKLING [Laz+19] has been introduced as an iterative scheme that optimizes for each k-space trajectory to be compliant with MRI hardware constraints (particularly maximum gradient and slew rate constraints), while ensuring that the overall sampling pattern obtained with all the trajectories follows a target sampling density. Further, the algorithm ensures that optimized k-space sampling pattern does not have any local clusters, leading to locally uniform sampling patterns. This algorithm was extended to 3D [Laz+20a; Cha+22d] and showed superior performance in both terms of a peaky point spread function and image quality.

However, a major drawback of SPARKLING algorithm is the need to setup a target sampling density as an input to the algorithm. In our earlier studies, we relied on heuristic methods to set this sampling density. The latter was parameterized to be radially decaying and its optimal parameters (decay, cutoff) were grid searched during retrospective reconstruction studies in which image quality was maximized as a function of optimized trajectories for varied target densities. However, this approach is too computationally expensive. Also, with a parameterized target density, the search space is too constrained, preventing us to obtain organ, imaging-contrast or orientation-specific sampling schemes. One way to tackle this problem is by learning the target sampling density using data-driven approaches.

In [Kno+11], the authors proposed a naive approach to choose the target sampling density by averaging the power spectra of multiple MR images in a dataset. This method results in sampling densities that enforce denser sampling in the low frequencies. In [Kno+11] the authors showed that this approach outperforms standard VDS and remains robust to variability in anatomy and orientation. However, this method focuses purely on the MRI dataset and is agnostic to the reconstruction technique. All MRI reconstruction algorithms enforce a prior (like sparsity in the wavelet or image gradient domain). Recent deep learning (DL) reconstruction algorithms [Kno+20; Muc+21b; RCS20] have learned more complex priors based on the organ or contrast-specific dataset. The target sampling density can be more efficient if it takes these priors into account and enforce denser samples in regions where the degree of uncertainty associated with such priors for reconstruction is higher.

More recently, methods like [She+20; Bah+20a] learn the sampling pattern for MRI in a data-driven manner while optimizing for image quality at the reconstruction stage. In the deep learning setting, LOUPE [Bah+20a] jointly optimizes the sampling density and the weights of a U-net architecture for image reconstruction. However, these studies are limited to Cartesian sampling. Most appealing contributions [Wei+20; Ved+20] tend to directly learn the trajectories in a data-driven manner under MR hardware constraints. Particularly, in [Wei+20], the authors use multi-resolution to overcome the problem of a large number of trainable parameters which crops up in such direct optimization. However, the final trajectories were similar to perturbed versions of the initialization.

In this work, we use the target density obtained by LOUPE as an input to the SPARKLING algorithm to generate 2D SPARKLING non-Cartesian trajectories. We carry out retrospective studies and compare them with those that result from other densities such as the average (log-)power spectra over the fastMRI dataset. We perform image reconstruction using both CS technique and the newly developed NC-PDNet [RSC21a] architecture which is a density compensated unrolled neural network for non-Cartesian MRI reconstruction. We conclude that the proposed solution (LOUPE+2D SPARKLING) outperforms other VDS approaches in terms of image quality.

6.2 Materials and methods

Here, we detail the methods used to optimize for sampling density and thereby how the latter is injected as an input to SPARKLING algorithm to generate non-Cartesian trajectories. We later study the performance of the corresponding sampling schemes on retrospective MR image reconstruction studies.

6.2.1 2D Non-Cartesian trajectories

Throughout this work, we follow the formulation we developed in [Cha+22d], for the case of 2D imaging. Let the MR image size be $N \times N$, over a field of view $\mathcal{F} \times \mathcal{F}$. Then the 2D k-space of the image is defined in $[-K_{\max}, K_{\max}]^2$, with $K_{\max} = \frac{N}{2\mathcal{F}}$. In all our trajectories, we kept $N = 320$ and $\mathcal{F} = 0.23$ m. For the sake of simplicity, let us normalize the k-space to $\Omega = [-1, 1]^2$. We are optimizing for the 2D k-space sampling pattern $\mathbf{K}$ which is composed of several shots N_c, $\mathbf{K} = (\mathbf{k}_i)_{i=1}^{N_c}$. Each 2D shot $\mathbf{k}_i(t) = (k_{i,x}(t), k_{i,y}(t))$ is controlled by the magnetic field gradients $\mathbf{G}_i(t) = (G_{i,x}(t), G_{i,y}(t))$ as follows: $\mathbf{k}_i(t) = \frac{\gamma}{2\pi} \int_0^t \mathbf{G}_i(\tau)\, d\tau$ with γ the gyro-magnetic ratio ($\gamma = 42.57$MHz/T for proton imaging). Each shot is sampled at the pace of gradient raster time Δt, throughout the readout time T_{obs}, resulting in $N_s = \lfloor \frac{T_{\text{obs}}}{\Delta t} \rfloor$ samples per shot. The k-space data from the scanner is sampled at dwell time δt, which in practice is a fraction of Δt. Thus the total received k-space samples are of the form $\mathcal{K} \in \mathbb{C}^{N_c \times N_s \times \frac{\Delta t}{\delta t}}$. In our studies, we used dwell time ($\delta t = 2\mu$s) and gradient raster time ($\Delta t = 10\mu$s), thereby having 5 times more k-space sample points than the measurements defined by the gradient wave forms.

The MR hardware constraints of maximum gradient strength ($G_{\max} = 40$ mT/m) and slew rate ($S_{\max} = 180$ T/m/s) results in a constrained trajectory with limited speed (α) and acceleration (β). Note that the speed constraint also handles the Nyquist sampling criterion (see [Cha+17]). We define this constraint set as $\mathcal{Q}_{\alpha,\beta}^{N_c}$, see [Cha+22d].

6.2.2 SPARKLING algorithm

Let the target sampling distribution be $\rho : \Omega \to \mathbb{R}$, with $\rho(x) \geq 0$ for all x and $\int \rho(x)\, dx = 1$. Given ρ, the SPARKLING algorithm optimizes for the k-space sampling pattern $\mathbf{K}$ such that the actual sampling distribution is closest to ρ, while being locally uniform. Although theoretically SPARKLING takes a continuous distribution ρ as input parameter, in practice, we discretize the distribution to obtain $\boldsymbol{\rho} \in \mathbb{R}^{N \times N}$. Further, the algorithm ensures that the each k-space shot $\mathbf{k}_i(t)$ in optimal $\widehat{\mathbf{K}}$ lies in $\mathcal{Q}_{\alpha,\beta}^{N_c}$. We can now summarize the SPARKLING algorithm as follows:

$$\widehat{\mathbf{K}} = \mathcal{S}(\boldsymbol{\rho}, \mathcal{Q}_{\alpha,\beta}^{N_c}, \mathbf{K_0}) \tag{6.1}$$

with $\mathbf{K_0}$ being the initialization. The detailed algorithm is presented in [Cha+22d]. Hereafter, we discuss different gridded distributions $\boldsymbol{\rho}$ that were obtained for our study.

6.2.3 Target sampling density learning

In this work, we broadly use four methods for estimating or learning a target sampling density. All these methods are data-driven and we rely on the fastMRI dataset [Zbo+18] to compute them. Let $\left\{\mathbf{x}_j \in \mathbb{R}^{N \times N}\right\}_{j=1}^{n}$ denote brain MR images from this dataset, where j is the scan

number and n is the total number of images (for simplicity, we used magnitude-only images). Let $\left\{\mathbf{v}_j \in \mathbb{C}^{N\times N}\right\}_{j=1}^{n}$ correspond to their respective discrete k-spaces on a grid (Fourier spectrum) obtained by a fast Fourier transform.

VDS-based

The first method we employed to obtain a density is based on naive VDS. For this, we parameterized the density as radially decaying with cutoff C and decay D as described in [Cha+22d]:

$$\rho_{vds}^{C,D}(x) = \begin{cases} \kappa & |x| < C \\ \kappa \left(\frac{C}{|x|}\right)^D & |x| > C \end{cases} \tag{6.2}$$

In our experiments, we heuristically grid searched for optimal parameters and used $C = 25\%$ and $D = 2$ as the best density.

Spectrum-based

Next we obtained a sampling density based on [Kno+11] which involves averaging the spectra of brain images from the fastMRI dataset. Let $\mathbf{v}_{avg}$ correspond to the average of all the spectra $\mathbf{v}_j$ in the dataset. Then we can normalize the 2D spectrum to obtain a sampling density $\boldsymbol{\rho}_{\text{sb}}$ on the $N \times N$ grid:

$$\boldsymbol{\rho}_{\text{sb}}(p,q) = \frac{\mathbf{v}_{avg}(p,q) - \min(\mathbf{v}_{avg})}{\sum_{p,q}\left[\mathbf{v}_{avg}(p,q) - \min(\mathbf{v}_{avg})\right]}\,. \tag{6.3}$$

Further, we observed that the spectra have very large magnitudes at lower frequencies as compared with higher frequencies. In an effort to flatten the distribution so that we may better balance all frequencies, we relied on an average *log-spectrum* $\mathbf{v}_{lavg}$ of the fastMRI images and obtained the distribution $\boldsymbol{\rho}_{\text{lsb}}$ by replacing $\mathbf{v}_{avg}$ with $\mathbf{v}_{lavg}$ in Eq. (6.3).

LOUPE-based

As the spectrum-based methods are agnostic to image reconstruction, to fill this gap we used the Cartesian acquisition model from LOUPE [Bah+20a]. LOUPE is actually a DL-based optimization scheme that learns a Cartesian under-sampling pattern for a prescribed sparsity level γ, which provides the percentage of discarded measurements as compared to a full sampling. Hence, γ is defined as the inverse of the under-sampling factor R $(= \frac{N\times N}{N_c \times N_s \times \frac{\Delta t}{\delta t}} = \frac{1}{\gamma}$ for non-Cartesian sampling). In practice, we used $R = 2.5$ $(\gamma = 0.4)$. Using LOUPE, we can learn a gridded sampling density $\boldsymbol{\rho}_{lb}$ by jointly optimizing the acquisition and reconstruction frameworks in the Cartesian domain. In [Bah+20a], the authors used conventional U-Net [RFB15a] for carrying out reconstruction. In contrast here, we integrate LOUPE's acquisition network with a modular cross-domain neural network called XPDNet [RSC21b] which stood second in the 2020 fastMRI brain reconstruction challenge [Muc+21b]. Hence, we jointly optimize for the sampling distribution $\boldsymbol{\rho}_{lb}$ and the reconstruction network. In regards to the LOUPE model, we initialize the *sigmoid sample slope* $s = 20$ and trained this network for 100 epochs over all the training set ($n = 4469$ MR images) and probed for the target sampling density. We ensured that there was no leaking of the k-space data into the reconstruction network by checking the resulting binary sampling masks (see [Bah+20a] for details).

6.2.4 Retrospective studies

With different target sampling distributions as input, we carried out an extensive retrospective study on 50 slices from the validation set of the FastMRI dataset. The k-space measurements were obtained by applying a forward NUFFT operator (F) to the input multi-coil brain MR images. We performed image reconstruction using two different methods:

CS reconstruction

First we used the the synthesis formulation of self-calibrating CS image reconstruction [El +18a] by solving for the wavelet coefficients $\mathbf{z}$ as follows:

$$\widehat{\mathbf{z}} = \underset{\mathbf{z}\in\mathbb{C}^{N\times N}}{\operatorname{argmin}} \frac{1}{2}\sum_{\ell=1}^{L} \|F_\Omega \mathbf{S}_\ell \mathbf{\Psi}^* \mathbf{z} - \mathbf{y}_\ell\|_2^2 + \lambda\|\mathbf{z}\|_1 \tag{6.4}$$

where the L is the number of coils. Here the data consistency is enforced with SENSE operators $(F_\Omega \mathbf{S}_\ell)_\ell$, where F_Ω is the NUFFT masked to Ω and $\mathbf{S}_\ell$ is sensitivity map for ℓ^{th} coil estimated by density compensated adjoint of the 20% of acquired k-space center (see details in [El +18a]). $\lambda > 0$ is the regularization parameter for ℓ_1-sparsity which was promoted in the wavelet domain $\mathbf{\Psi}$. For our reconstructions, we used Symlet 8 wavelet with 4 scales for $\mathbf{\Psi}$. The regularization parameter λ was grid searched between $(10^{-4}, 10^0)$ while maximizing for the reconstruction quality using structural similarity index (SSIM) in retrospective reconstruction. In order to accelerate convergence, we preconditioned the k-space using density compensation. The compensation weights were estimated with 10 iterations of method as described in [PM99]. Final MR images were reconstructed as $\widehat{\mathbf{x}} = \mathbf{\Psi}\widehat{\mathbf{z}}$.

DL reconstruction network (NC-PDNet)

For an extension into DL-based reconstruction, we used NC-PDNet [RSC21a], which is a non-Cartesian extension of the XPDNet used for learning the sampling density. More precisely, we used a density compensated unrolled non-Cartesian reconstruction network, whose parameters are the same as those described in [RSC21a]. This model was trained for 70k gradient descent steps on the respective contrasts (T_1-w and T_2-w) from multi-coil brain dataset with SPARKLING trajectories obtained in Fig. 6.1.

6.3 Results

In this section we briefly present the densities and trajectories for various methods of estimating the target sampling densities as described in Sec. 6.2.3. Then we briefly go through the retrospective reconstruction results that we obtained.

6.3.1 Densities and trajectories

The varied target sampling densities and their respective SPARKLING trajectories are presented in Fig. 6.1. We see that the direct spectrum-based density $\boldsymbol{\rho}_{sb}$ is extremely dense at the center of k-space, leading to really dense sampling here in the respective trajectories. The *log-spectrum* method does indeed flatten out the density $\boldsymbol{\rho}_{lsb}$, allowing the trajectories to explore more high frequencies. Finally, the LOUPE based density does oversample the center of k-space resulting in a scheme very similar to variable density sampling. However, the density $\boldsymbol{\rho}_{lb}$ from LOUPE is more grainy since the learning of this density happens on a Cartesian grid.

6.3.2 Retrospective image reconstruction studies

Quantitative results

We carried out retrospective studies on 50 slices of the validation data (two imaging contrasts, namely T_1 and T_2) in the fastMRI dataset for all the above generated trajectories. We computed the SSIM and peak signal-to-noise ratio (PSNR) metrics on the reconstructed MR images with a mask on the brain in order to assess image quality. We present the results as boxplots and annotated the significance as paired t-test in Fig. 6.2.

Firstly, we note that all methods perform pretty decently as long as the sampling density has been optimized, with NC-PDNet consistently outperforming traditional reconstruction schemes. However, we see that the SPARKLING trajectories with $\boldsymbol{\rho}_{lb}$ densities consistently

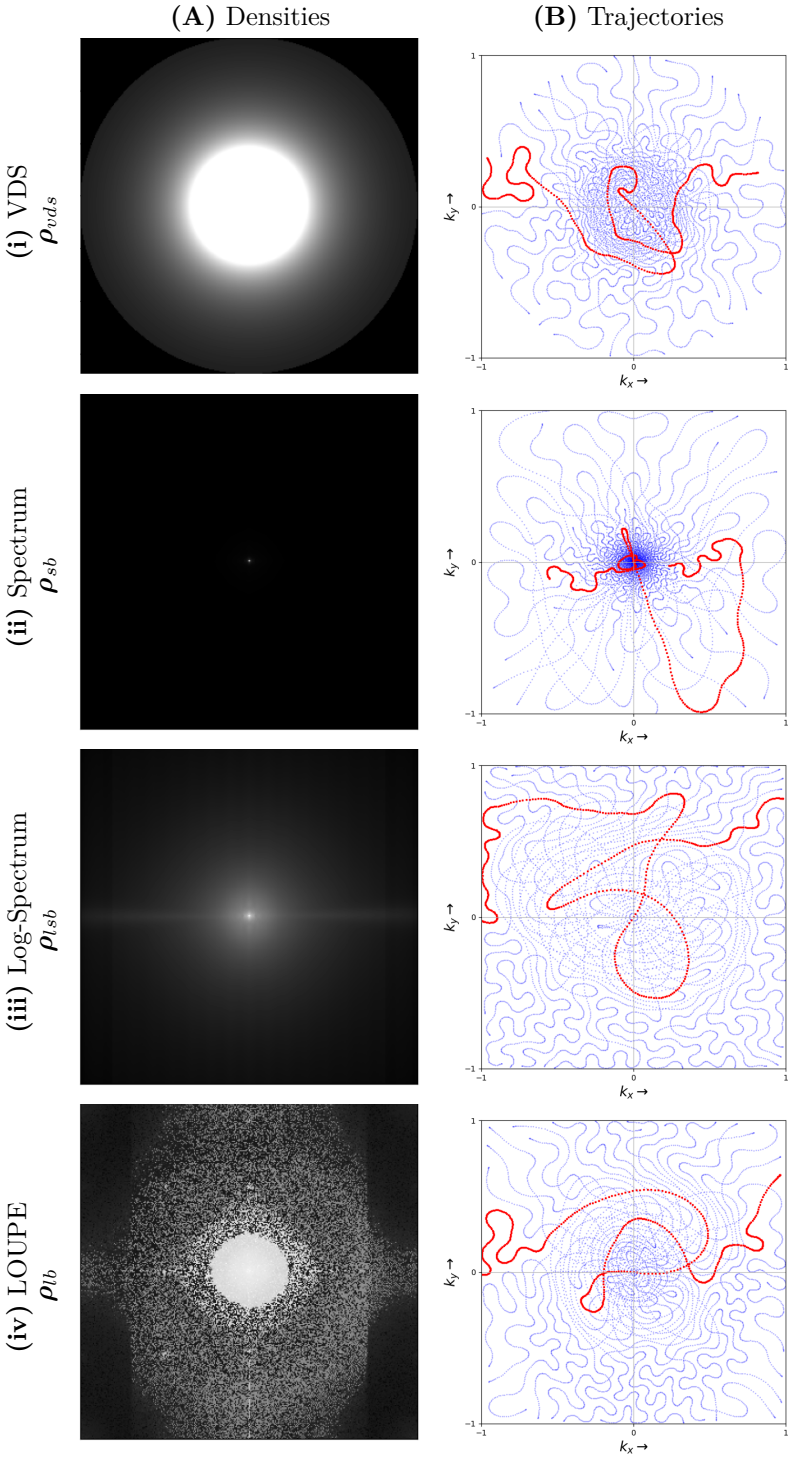


Figure 6.1 – **(A):** The target sampling densities obtained for T_1-weighted images with: **(i)** VDS ($\boldsymbol{\rho}_{vds}$), a radially decaying parameterized density, with C=25% and D=2 in [Cha+22d]; **(ii)** Average spectrum ($\boldsymbol{\rho}_{sb}$) over the dataset based on [Kno+11]; **(iii)** Average logarithm of the spectrum ($\boldsymbol{\rho}_{lsb}$) over the dataset, to flatten the density in **(ii)**; **(iv)** LOUPE ($\boldsymbol{\rho}_{lb}$) [Bah+20a] coupled with XPDNet [RSC21b] reconstruction. **(B):** Corresponding k-space trajectories generated with $N_c = 16$ ($R = 2.5$), $N_s = 512$, $G_{\max} = 40$ mT/m and $S_{\max} = 180$ T/m/s. For illustration purpose, a single shot is colored in red.

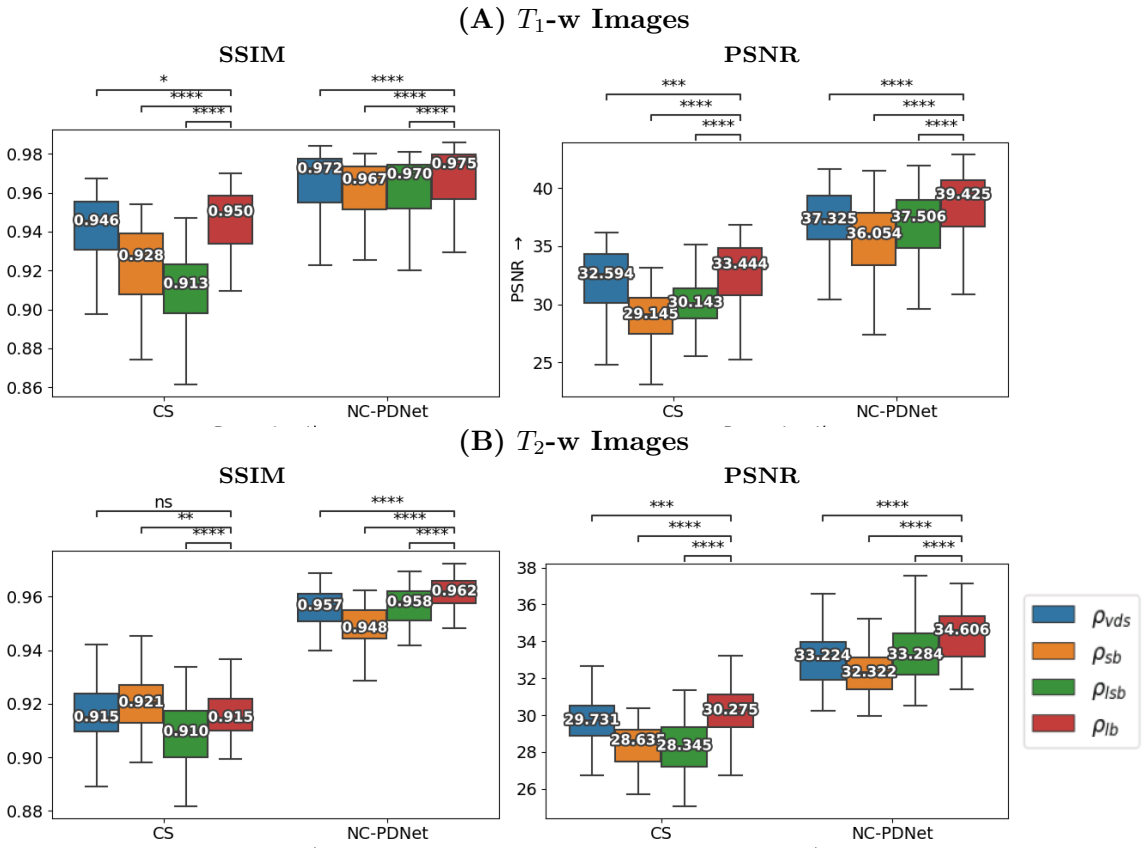


Figure 6.2 – Retrospective study on different trajectories for $R = 2.5$ on 50 slices of **(A)** T_1-w and **(B)** T_2-w Images. The reconstructions were performed with both CS based reconstruction (Sec. 6.2.4) and using NC-PDNet (Sec.6.2.4) The median SSIM and PSNR scores are indicated.

perform well throughout with SSIMs always larger than 0.95/0.91 for T1/T2 contrast (red boxes). Additionally, this method has the highest PSNR. This confirms our hypothesis that a method which is both data-driven and reconstruction aware outperforms its competitors for learning a target sampling density. Finally, we noticed that $\boldsymbol{\rho}_{vds}$ performs similarly to $\boldsymbol{\rho}_{lb}$ with respect to SSIM in most cases. This might be due to the properties of k-space content in brain imaging, which is radially symmetric. Hence optimizing for a radially decaying density gives similar performances to LOUPE-based methods. However, it is worth mentioning that this optimization of parameterized density is very computationally intensive as it involves both trajectory generation and retrospective reconstruction in order to understand which parameter affects the most image quality.

Qualitative results

For visual inspection, we present the results of image reconstruction from data undersampled using SPARKLING trajectories generated for various target densities in Fig. 6.3 (T_1-w images) and Fig. 6.4 (T_2-w images). For the sake of space, we only report the best reconstruction results, i.e. with NC-PDNet. For T_1-weighted contrast, we show that all methods give similarly performing results, however $\boldsymbol{\rho}_{vds}$ and $\boldsymbol{\rho}_{lb}$ provide the best SSIM scores. Further, we observe that in this case, $\boldsymbol{\rho}_{vds}$ is slightly better than $\boldsymbol{\rho}_{lb}$. On the contrary, for T_2-w contrast, $\boldsymbol{\rho}_{lb}$ outperforms the other densities as reflected both visually in Fig. 6.4 and quantitatively (see Fig. 6.2).

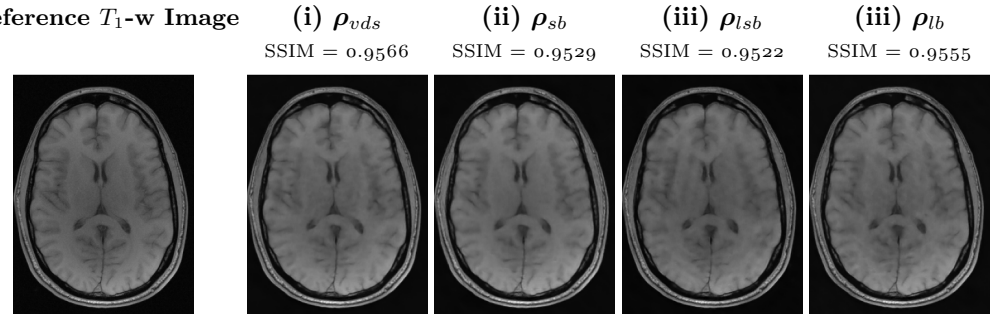


Figure 6.3 – NC-PDNet-based image reconstruction for retrospective $T1$-w imaging with slice 6 in *file_brain_AXT1_201_6002725.h5* from validation data in fastMRI dataset for different target sampling densities.

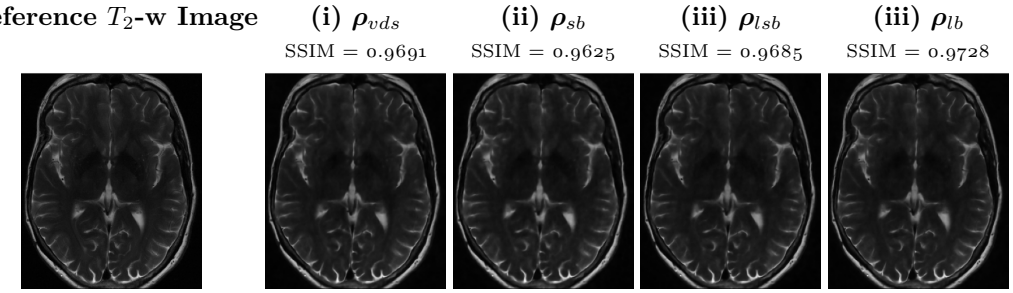


Figure 6.4 – NC-PDNet-based image reconstruction for retrospective $T2$-w imaging with slice 5 in *file_brain_AXT2_200_2000019.h5* from validation data in fastMRI dataset for different target sampling densities.

6.4 Conclusions

In this study, we addressed the main drawback of the SPARKLING algorithm, namely the need for a good target sampling density as an input parameter. We setup four different methods to generate optimized target sampling densities and design SPARKLING trajectories accordingly. We showed that the LOUPE-based approach is the most promising as it provides consistent results across contrasts. A limitation of this work is that there remains some split between the acquisition and reconstruction models in a fully non-Cartesian setting. Under the current study, the sampling density was jointly optimized with a Cartesian DL reconstruction network. Then, non-Cartesian SPARKLING trajectories were generated and retrospective validation was performed using a non-Cartesian DL network. There is thus still a gap between the training and validation stage in this pipeline. In spite of this limitation, we obtained promising results. In terms of perspective, we plan to work on a joint network between NC-PDNet and SPARKLING to efficiently learn the k-space trajectories in a data-driven manner, under the MR Hardware constraints.

* * *
* *
*

Chapter 7

Learning trajectories with reconstructor

This chapter covers content that has been published in a peer-reviewed journal:

- **Chaithya, G R** and P. Ciuciu. "Jointly learning Non-Cartesian k-space trajectories and reconstruction networks for 2D and 3D MR imaging through projection". In: *special issue on AI in MRI: Frontiers and Applications, Bioengineering*

A part of this work was also presented in an international peer-reviewed conference with proceedings:

- **Chaithya, G R** and P. Ciuciu. "Benchmarking learned non-Cartesian k-space trajectories and reconstruction networks". In: *ISMRM*. 3308. London, UK, May 2022

Compressed sensing in Magnetic resonance Imaging essentially involves the optimization of 1) the sampling pattern in k-space under MR hardware constraints and 2) image reconstruction from undersampled k-space data. Recently, deep learning methods have allowed the community to address both problems simultaneously, especially in the non-Cartesian acquisition setting. This work aims to contribute to this field by tackling some major concerns in existing approaches. Particularly, current state-of-the-art learning methods seek hardware compliant k-space sampling trajectories by enforcing the hardware constraints

through additional penalty terms in the training loss. Through ablation studies, we rather show the benefit of using a projection step to enforce these constraints and demonstrate that the resulting k-space trajectories are more flexible under a projection-based scheme, which results in superior performance in reconstructed image quality. In 2D studies, our novel PROjection for Jointly lEarning non-Cartesian Trajectories while Optimizing Reconstructor (PROJeCTOR) trajectories presents an improved image reconstruction quality at 20-fold acceleration factor on the fastMRI data set with SSIM scores of nearly 0.92-0.95 in our retrospective studies as compared to corresponding Cartesian reference and also see 3-4dB gain in PSNR as compared to earlier state-of-the-art methods. Finally we extend the algorithm to 3D and by comparing optimization as learning based projection schemes, we show that data-driven joint learning based PROJeCTOR trajectories outperform model-based methods like SPARKLING through a 2dB gain in PSNR and 0.02 gain in SSIM.

7.1 Introduction

A major challenge limiting the use of Magnetic Resonance Imaging (MRI) is long acquisition times, arising due to short decay of the MR signal which is used to sample multi-dimensional k-space data through numerous and repetitive radio-frequency pulses. Using Compressed Sensing (CS) theories [Lus+07], significant speed up can be obtained by undersampling the k-space according to Variable Density Sampling (VDS) [Puy+11; Cha+13; Cha+14; Adc+13; Boy+19], whose shape depends on the underlying anatomy, contrast and coil structure. Non-Cartesian sampling can be used to efficiently achieve VDS of k-space, as this type of sampling, which relies on curves, is more flexible and efficient compared to straight lines used in traditional Cartesian acquisitions. While conventional non-Cartesian sampling patterns like spiral, radial, rosette, etc. [Ahn+86b; Mey+92; Jac+92; Nol97; Law+09; Lus+05a] have been proposed in literature which can sample the k-space according to VDS, they do not sample at a well defined user specified Target Sampling Density (TSD). Tailoring such non-Cartesian trajectories according to a MR imaging protocol and a given TSD is hard as these k-space sampling curves or trajectories are constrained by the MR hardware limits notably on the maximum gradient magnitude $G_{\max}$ and slew rate $S_{\max}$.

To meet these constraints in a safe manner, the SPARKLING was introduced in [Laz+19; Laz+20a] and then extended to 3D [Cha+22d] as an iterative procedure to optimize a non-Cartesian k-space sampling pattern according to a prescribed TSD. Such patterns are typically segmented in multiple shots or k-space trajectories, each of them being compliant with the above mentioned MR hardware constraints. Further, the algorithm results in locally uniform sampling patterns and thus avoids holes and clusters in k-space. However, SPARKLING is a model-driven framework, which is characterized by a TSD that needs to be known in advance to feed the optimization process. Previously in [CRC21], to address this issue, we learned the TSD in a deep learning setting using LOUPE [Bah+20a] as an acquisition model. Although this allowed us to improve reconstruction performances, there was still a mismatch in the learning process. Using LOUPE [Bah+20a], gridded TSD was learned in the Cartesian domain, while the actual trajectory being optimized was non-Cartesian. Additionally, we had to learn a different non-Cartesian image reconstruction model (e.g. a convolutional neural network or CNN) that was disconnected from the optimized trajectories, making the overall process computationally expensive. Further, as such disjointedness between training a TSD and testing on different non-Cartesian trajectories and image reconstruction neural nets could lead to suboptimal results, there is a need to jointly learn both the TSD and the image reconstruction deep learning architecture in a non-Cartesian setting.

Recently, new methods [Wei+20; Wan+21; Ved+20] have been developed to overcome the need for estimating a TSD, through direct joint learning of the non-Cartesian k-space sampling trajectories and MR image reconstruction in a data-driven manner on the fastMRI dataset [Zbo+18]. In [Wei+20; Ved+20], the authors jointly learned Physics-informed learned optimal trajectories (PILOT) trajectories along with U-net parameters as a reconstruction model to denoise the basic image yielded by the adjoint of the Nonuniform Fast Fourier Transform (NUFFT) operator. However, this method relies on auto-differentiation of the NUFFT operator, which is inaccurate numerically as observed in [WF23], resulting in sub-

optimal local minima. This suboptimality was actually reflected in the final shape of the learned trajectories, which only slightly deviated from their initialization.

In B-spline parameterized Joint Optimization of Reconstruction and K-space trajectories (BJORK) [Wan+21], the authors use [WF23] to obtain a more accurate Jacobian approximation of the NUFFT operator. Both above referenced approaches [Wei+20; Wan+21] enforced the hardware constraints by adding penalty terms to the the loss that is minimized during training. Although a viable option, this requires tuning a hyper-parameter associated with each of these penalty terms in the cost function. Moreover it does not guarantee that the optimized trajectories will strictly meet these constraints. Further, these penalty terms affect the overall gradients of the loss function, thereby resulting in suboptimality of the trajectories. In BJORK [Wan+21] the trajectories were parameterized with B-spline curves in order to reduce the number of trainable parameters. Although this strategy drastically minimizes the search space and the training time, such parameterization severely limits the degrees of freedom of the trajectories and prevents them from an improved exploration of the k-space. Finally, both methods do not make use of Data Consistency (DC) which plays a key role in obtaining clearer MR images in the non-Cartesian deep learning setting [Ram+22b].

In this work, we first develop a generic model for PROJeCTOR. More precisely, we introduce a method that learns the k-space trajectories in a data-driven manner while embedding a projected gradient descent algorithm [Cha+16] to fulfill the hardware constraints during the training stage. Unlike BJORK, we directly learn the k-space sampling trajectories and use multi-resolution [Leb+19] similar to SPARKLING to limit the number of trainable parameters at each step. Then, we compare these PROJeCTOR results to two state-of-the-art methods, PILOT [Wei+20] and BJORK [Wan+21] in 2D MRI. In a more controlled setting we show the importance of the projection step during the optimization of k-space trajectories and demonstrate its superiority over penalty-based methods like PILOT and BJORK to enforce hardware constraints. Finally, we compare and show the superiority of data-driven PROJeCTOR trajectories compared to model-based non-Cartesian SPARKLING trajectories.

7.2 Materials and Methods

In this section, we present a generic and modular framework (Fig. 7.1) for learning non-Cartesian k-space trajectories and deep neural networks for MR image reconstruction. Particularly, we discuss 2 sub-models namely, 1) an Acquisition model parameterized by k-space trajectory and 2) an Reconstruction model parameterized by a deep neural network. Later, we discuss in detail how to handle the MR hardware constraints and which approach seems the most efficient within the sampling pattern optimization process to end up with hardware compliant k-space trajectories.

7.2.1 Data and preprocessing

In order to reduce the memory footprint and the training time, we did not process multicoil k-space data as input in the pipeline shown in Fig. 7.1. Instead we learn the trajectories and image reconstruction model on emulated single coil data obtained using virtual coil combination [Par+14] of per-channel images. This is done through phase reconstruction from multi-coil data through the use of a virtual reference coil. This virtual-reference coil is generated as a weighted combination of measurements from all receiver coils. The multiple phase-corrected coil complex images are combined using the inverse covariance matrix, to result in a complex image with optimal estimates of the absolute magnetization phase (see [Par+14] for mathematical details).

Overall, we rely on notations developed in [Cha+22d], and we assume isotropic resolution and FOV with matrix size in each axis as N. This assumption is purely for notational convenience and does not limit the applicability of our framework to isotropic data. If D is the imaging dimension, we denote an MR image or volume as $\mathbf{x} \in \mathbb{C}^{N^D}$, over a field of view $\mathcal{F}^D$. Throughout the manuscript, we refer to $\mathbf{x}$ as MR image, while it can be MR volume when $D = 3$. The k-space of this acquisition is defined in $[-K_{\max}, K_{\max}]^D$, with $K_{\max} = \frac{N}{2\mathcal{F}}$. However, for the sake of simplicity, we normalize the k-space to $\Omega = [-1, 1]^D$. For both

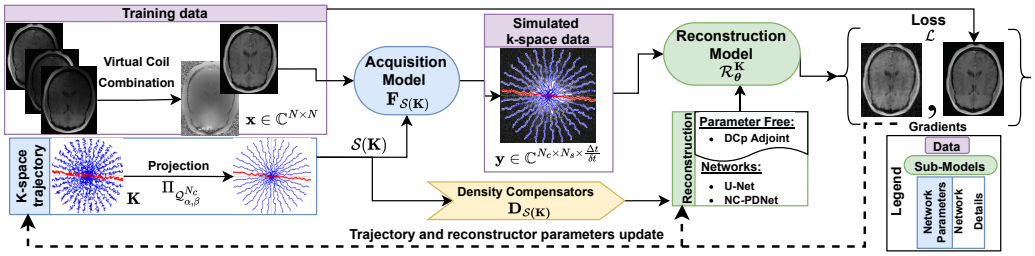


Figure 7.1 – A generic learning-based framework for joint optimization of the MRI acquisition and reconstruction models. This framework consists of two sub-models: 1) The Acquisition model $\mathbf{F}_{\mathcal{S}(\mathbf{K})}$ parameterized by the k-space sampling trajectories $\mathbf{K}$, interpolated through linear interpolation $\mathcal{S}(\mathbf{K})$ and 2) The reconstruction model $\mathcal{R}^{\mathbf{K}}_{\theta}$ parameterized by θ. The input training data consists of emulated single coil complex images, from which simulated k-space data is obtained through $\mathbf{F}_{\mathcal{S}(\mathbf{K})}$. The loss $\mathcal{L}$ is calculated between the reconstructed image and the ground truth. The gradients are backpropagated to result in k-space trajectory and reconstructor parameters update. Projection $\Pi_{\mathcal{Q}^{N_c}_{\alpha,\beta}}$ is carried out after trajectory update to make sure it satisfies the hardware constraints and lies in the constraint set $\mathcal{Q}^{N_c}_{\alpha,\beta}$. Further, the density compensators $\mathbf{D}_{\mathcal{S}(\mathbf{K})}$ of the k-space trajectory serves as input to the reconstruction network.

2D and 3D imaging, we take Observation time (T_{Obs})=5.12ms (readout time), giving us $N_s = 512$ samples per trajectory (see details in subsection 7.2.2). This readout value is fully compatible with those used in T1 and T2-weighted imaging.

For our experiments in 2D imaging, we used the fastMRI brain MR data set [Zbo+18], which consists of 1447 T_1 and 2678 T_2-weighted images with $N = 320$. In contrast, for validation in 3D imaging, we used the Calgary brain data set [Sou+18], which consists of 167 T_1-w MR volumes at 1 mm isotropic sagittal acquisitions, with matrix size $256 \times 224 \times 170$. For both imaging protocols, we used an Acceleration Factor (AF) $= \frac{N^{D-1}}{N_c}$ of 20 (see [Cha+22d]), resulting in a number of trajectories $N_c = 16$ for 2D imaging and $N_c = 1681$ for 3D imaging (see details in subsection 7.2.2).

7.2.2 *K*-space trajectory (K)

The acquisition model is parameterized by a k-space sampling pattern $\mathbf{K}$ which is composed of N_c shots, $\mathbf{K} = (\mathbf{k}_i)_{i=1}^{N_c}$. Each shot can be played by the scanner hardware at the pace of gradient raster time Δt, throughout the readout time T_{obs}, resulting in $N_s = \lfloor \frac{T_{\text{obs}}}{\Delta t} \rfloor$ samples per shot and overall sampling pattern as $\mathbf{K} \in \Omega^{N_c \times N_s}$.

The k-space trajectories are constrained in speed and acceleration by the maximum gradient strength $G_{\max}$ and maximum slew rate $S_{\max}$, respectively. Additionally, affine constraints are added to the trajectory design to ensure that the center of k-space is sampled at Echo Time (TE) in every shot, resulting in stable and required target contrast of reconstructed MR images. From [Cha+16; Cha+22d], we model these constraints as follows:

$$\mathcal{Q}^{N_c}_{\alpha,\beta} = \left\{ \begin{array}{c} \forall i = 1, \dots, N_c, \quad \mathbf{k}_i \in \mathbb{R}^{3 \times N_s}, \\ \mathbf{A}\mathbf{k}_i = \mathbf{v}, \\ \|\mathbf{k}_i\|_\infty \leq 1, \ \|\dot{\mathbf{k}}_i\|_{2,\infty} \leq \alpha, \ \|\ddot{\mathbf{k}}_i\|_{2,\infty} \leq \beta, \end{array} \right\} \tag{7.1}$$

where

$$\dot{\mathbf{k}}_i[n] = \frac{\mathbf{k}_i[n] - \mathbf{k}_i[n-1]}{\Delta t}$$
$$\ddot{\mathbf{k}}_i[n] = \frac{\mathbf{k}_i[n+1] - 2\mathbf{k}_i[n] + \mathbf{k}_i[n-1]}{\Delta t^2}$$
$$\|\mathbf{c}\|_{2,\infty} = \sup_{0 \le n \le N_s - 1} \left(|c_x[n]|^2 + |c_y[n]|^2 + |c_z[n]|^2\right)^{1/2},$$

for all $\mathbf{c} \in \Omega^{N_s}$ and (α, β) are obtained by normalizing hardware and Nyquist constraints to the sampling domain Ω (from [Laz+19]):

$$\alpha = \frac{1}{K_{\max}} \min\left(\frac{\gamma G_{\max}}{2\pi}, \frac{1}{FOV \cdot \delta t}\right) \tag{7.2a}$$
$$\beta = \frac{\gamma S_{\max}}{2\pi K_{\max}} \tag{7.2b}$$

The TE point constraints are modeled through $\mathbf{A}$ and $\mathbf{v}$ in (7.1) (see [Cha+16] for details and more complex affine constraints). $\mathbf{A}$ and $\mathbf{v}$ are tailored to have the following equivalent expression on each k-space trajectory:

$$\mathbf{k}_i^d[k_{\text{te}}] = 0 \qquad \begin{array}{c} \forall\, i \in \{1, \dots N_c\}, \\ \forall\, d \in \{x, y, z\}, \\ k_{\text{te}} = \lfloor \frac{TE}{\Delta t} \rceil\,. \end{array} \tag{7.3}$$

7.2.3 Acquisition Model ($\mathbf{F_K}$)

With the k-space sampling pattern $\mathbf{K}$, we model the acquisition process at the MR scanner with non-uniform fast Fourier transform (NUFFT) [FS03] operator $\mathbf{F_K}$. However, in practice, the k-space data is sampled in Analog to Digital Converter (ADC) at every dwell time δt, with $o = \frac{\Delta t}{\delta t} \geq 1$ the oversampling factor along each trajectory. Thus, a more realistic acquisition model of k-space data $\mathbf{y} \in \mathbb{C}^{N_c \times N_s \times o}$ is:

$$\mathbf{y} = \mathbf{F}_{\mathcal{S}(\mathbf{K})}\mathbf{x} + \epsilon \tag{7.4}$$

where $\mathcal{S}$ is linear interpolator, which interpolates the k-space trajectory to have $o \times N_s$ samples during readout, to model the oversampling by ADC and ϵ is the simulated noise which is already present in the data set as they are prospectively acquired by the MR system.

As the k-space trajectories are non-Cartesian, this creates a variable density sampling in k-space, due to which a simple adjoint of NUFFT operator $\mathbf{F}_{\mathbf{K}}^*$ is not close to the inverse operator and is not sufficiently accurate to reconstruct a clear MR image. To prevent this, a density compensation (DC) mechanism has been introduced in the non-Cartesian image reconstruction community for more than 20 years [PM99]. It allows us to more fairly balance the weights of k-space samples associated with the low and high frequencies during the iterative reconstruction process. Following this principle, we obtained $\mathbf{D}_{\mathcal{S}(\mathbf{K})}$ for the linearly interpolated k-space trajectory $\mathcal{S}(K)$, which is computed by 10 iterations of the algorithm described in [PM99]. As noted in [Ram+22b], DC is crucial for deep learning based reconstruction to avoid numerical issues and result in better reconstructed image quality.

7.2.4 Reconstruction model: Deep neural network ($\mathcal{R}_{\mathbf{K}}^{\theta}$)

The reconstruction network $\mathcal{R}_{\mathbf{K}}^{\theta}$ is a deep neural network that reconstructs an MR image $\widehat{\mathbf{x}}$ from the k-space data $\mathbf{y}$ and the k-space trajectory $\mathcal{S}(\mathbf{K})$. The estimated DC are also provided as input to the network, to better condition the reconstruction problem resulting in faster convergence giving us:

$$\widehat{\mathbf{x}} = \mathcal{R}_{\mathbf{K}}^{\theta}(\mathbf{y}, \mathbf{D}_{\mathcal{S}(\mathbf{K})})\,. \tag{7.5}$$

A simple parameter-free reconstruction would be the density compensated adjoint, i.e. $\mathcal{R}^{\theta}_{\mathbf{K}} = \mathbf{D}_{\mathcal{S}(\mathbf{K})}\mathbf{F}^{*}_{\mathbf{K}}$. To go further, we implemented the density compensated non-Cartesian primal dual network (NC-PDNet [Ram+22b]) as the reconstruction network. The latter alternates between a data consistency step in k-space and convolutional neural network (CNN) based denoising in the image domain with kernel size 3×3 in 2D and $3 \times 3 \times 3$ in 3D. We used the same network architecture as in [Ram+22b] except that this time we expanded the architecture over 12 unrolled iterations and the number of filters per iteration $N_f = 32$ filters.

7.2.5 Loss, gradients and optimizer

The reconstruction error used as loss function $\mathcal{L}_r$ (between the reference MR image $\mathbf{x}$ and its reconstruction $\widehat{\mathbf{x}}$) in this study was inspired by [Pez+20a] and is defined as a weighted sum of ℓ_1, ℓ_2 and multi-scale structural similarity index (S) [WSB03a]:

$$\mathcal{L}_r(\mathbf{x}, \widehat{\mathbf{x}}) = \alpha(1 - S(\mathbf{x}, \widehat{\mathbf{x}})) + \bar{\alpha}||\mathbf{x} - \widehat{\mathbf{x}}||_1 + \frac{\bar{\alpha}^2}{2}||\mathbf{x} - \widehat{\mathbf{x}}||_2$$

with $\bar{\alpha} = 1 - \alpha$ and the value of α was tuned to 0.995 to give nearly equally balanced loss terms. The training was carried out by minimizing reconstruction loss $\mathcal{L}_r$ with respect to both parameters θ of the reconstruction network and k-space trajectory $\mathbf{K}$ as follows:

$$(\widehat{\mathbf{K}}, \widehat{\boldsymbol{\theta}}) = \underset{(\mathbf{K} \in \mathcal{Q}^{N_c}_{\alpha,\beta}, \boldsymbol{\theta})}{\arg\min} \mathcal{L}_r\left(\mathbf{x}, \mathcal{R}^{\boldsymbol{\theta}}_{\mathbf{K}}\left(\mathbf{F}_{\mathcal{S}(\mathbf{K})}\mathbf{x}\right)\right) \tag{7.6}$$

For optimizing the trajectory $\mathbf{K}$, we derived the gradient of the loss function $\mathcal{L}_r$ with respect to $\mathbf{K}$:

$$\frac{\partial \mathcal{L}_r(\mathbf{x}, \widehat{\mathbf{x}})}{\partial \mathbf{K}} = \nabla \mathcal{L}_r(\mathbf{x}, \widehat{\mathbf{x}}) \frac{\partial \widehat{\mathbf{x}}}{\partial \mathbf{K}} = \nabla \mathcal{L}_r(\mathbf{x}, \widehat{\mathbf{x}}) \frac{\partial \mathcal{R}^{\boldsymbol{\theta}}_{\mathbf{K}}(\mathbf{y})}{\partial \mathbf{K}} \tag{7.7}$$

For ease of mathematical derivation, here we take the case of a parameter-free reconstruction as described in Sec. 7.2.4 with $\mathcal{R}^{\theta}_{\mathbf{K}} = \mathbf{D}_{\mathcal{S}(\mathbf{K})}\mathbf{F}^{*}_{\mathcal{S}}(\mathbf{K})$. In order to simplify this gradient calculation and reduce its computational complexity, we neglect the contribution of gradients from density compensators $D_{\mathcal{S}(\mathbf{K})}$. This contribution of gradients from $D_{\mathcal{S}(\mathbf{K})}$ was also ignored in realistic implementations to reduce gradient computation time and GPU memory requirements. These assumptions lead to the following gradient expression:

$$\frac{\partial \mathcal{L}_r}{\partial \mathbf{K}} = \nabla \mathcal{L}_r \left(\frac{\partial \widehat{\mathbf{x}}}{\partial \mathbf{D}_{\mathcal{S}(\mathbf{K})}\mathbf{y}} \mathbf{D}_{\mathcal{S}(\mathbf{K})} \frac{\partial \left(\mathbf{F}_{\mathcal{S}(\mathbf{K})}\mathbf{x}\right)}{\partial \mathbf{K}} + \frac{\partial \mathbf{F}^{*}_{\mathcal{S}(\mathbf{K})}}{\partial \mathbf{K}} \right)$$

In order to compute the gradient of NUFFT operators $\mathbf{F}_{\mathcal{S}(\mathbf{K})}$ and $\mathbf{F}^{*}_{\mathcal{S}(\mathbf{K})}$ with respect to the k-space trajectory $\mathbf{K}$, we used [WF23] to obtain a fast and accurate approximation of the Jacobians. As these underlying gradients vary extremely in norm depending on the k-space region (as noted in [Gou+21]), we used the ADAM optimizer for learning the trajectories, while we relied on a rectified-ADAM solver for optimizing the image reconstruction network $\mathcal{R}^{\mathbf{K}}_{\theta}$.

During training, the gradient descent was carried out stochastically with a batch size of 64 in 2D, while due to memory limitations, it was limited to 1 in 3D. However, as the gradients with respect to k-space trajectory were extremely noisy for this low batch size in 3D, we relied on a smaller learning rate of 2×10^{-4} as compared to 10^{-3} in 2D runs. On the other hand, for the optimization for the reconstruction networks, the corresponding gradients were more reliable and hence the learning rate was always set to 10^{-3}. The noise levels in gradients and their reliablity are quantified through the descent rate of the loss while optimizing with a fixed learning rate of 10^{-3} at varying batch sizes obtained through gradient accumulation. During gradient accumulation, gradients for the target batch size was obtained by running the network sequentially on multiple single data points repeatedly and accumulating the gradients.

7.2.6 Multi-Resolution

Inspired by SPARKLING [Cha+22d], the learning of the k-space sampling trajectories was performed using a multi-resolution strategy [Leb+19] which starts by learning $2^{R^{\max}}$ times decimated sampling trajectories $\mathbf{K}$ at the maximal $R^{\max} = 5$ decimation level. Next, the solution $\widehat{\mathbf{K}}^{2^{R^{\max}}}$ at the resolution level $R^{\max}$ was then interpolated and used as a warm restart for the up-sampled problem at resolution level $R^{\max} - 1$.

We used dyadic scaling and trained our trajectory over five decimation levels ($R^{\max} = 5$). This implies that the underlying trajectories were optimized first at $2^5 = 32$ decimation level (32 times downsampled trajectory), followed by upscaling the problem by 2, following the decimation levels as $16 \to 8 \to 4 \to 2 \to 1$. This multiresolution strategy was instrumental in ensuring fast convergence toward a local minimizer. Indeed initially the optimization is carried out with faster convergence as we coarsely optimize the k-space trajectory over a reduced number of locations ($R = R^{\max} = 5$). Then the process is refined at higher resolutions as we approach convergence ($R = 1$).

7.2.7 Constraints: Projection vs Penalty

A common method in the literature [Wan+21; Wei+20; Ved+20; Wan+22] to enforce these constraints is to add a penalty $\mathcal{L}_\rfloor(\mathbf{K})$ to the loss $\mathcal{L}$, which acts like a regularizer on the k-space trajectories $\mathbf{K}$ being optimized. With this, the loss function $\mathcal{L}$ becomes:

$$\mathcal{L}(\mathbf{x}, \widehat{\mathbf{x}}, \mathbf{K}) = \mathcal{L}_r(\mathbf{x}, \widehat{\mathbf{x}}) + \mathcal{L}_c(\mathbf{K}), \tag{7.8}$$

where the penalty $\mathcal{L}_c(\mathbf{K})$ follows the expression from [Wan+21; Wan+22]:

$$\mathcal{L}_c(\mathbf{K}) = \sum_{i=1}^{N_c} \sum_{n=1}^{N_s} \left(\lambda_1 \phi_\alpha \left(\|\dot{\mathbf{k}}_i[n]\|_2\right) + \lambda_2 \phi_\beta \|\ddot{\mathbf{k}}_i[n]\|_2\right) + \lambda_3 \phi_0 \|\mathbf{k}[k_{\mathrm{TE}}]\|_2 \tag{7.9}$$

with $\phi_a(x) = \max(0, x - a)$ and λ_1, λ_2 and λ_3 are hyper-parameters to balance the penalty terms with respect to the reconstruction loss $\mathcal{L}_r$.

However, this penalty based approach has the following limitations:

- **Need for hyper-parameter tuning:** Under the penalty based formulation, the hyper-parameters $\lambda_i \; \forall \, i \in \{1, 2, 3\}$ need to be tuned, which requires additional computation. Note that while we can view Eq. (7.8) as an augmented Lagrangian form for the constrained optimization problem Eq. (7.6), the corresponding Karush-Kuhn-Tucker (KKT) conditions are computationally complex to be solved. Further, as we do not satisfy the Slater's conditions, as the reconstruction loss $\mathcal{L}_r$ is non-convex, the solutions of the KKT conditions are not guaranteed to be global minima.

- **Influence of gradients and convergence:** With the addition of penalty terms $\mathcal{L}_c$, the gradient updates involve added gradients from these penalties $\nabla \mathcal{L}_c$, which influence the overall trajectory development and hence the final optimized k-space trajectories. Gradient updates with these additional gradient terms can no longer guarantee optimal image reconstruction by minimizing the reconstruction loss $\mathcal{L}_r$.

- **Guarantee of admissibility:** Finally, the optimization of the augmented Lagrangian form does not guarantee that the final optimized k-space trajectory $\mathbf{K}$ satisfies the constraints Eq. (7.1).

To tackle the above issues, we implemented the projector $\Pi_{\mathcal{Q}_{\alpha,\beta}^{N_c}}$ from [Cha+16] to project the k-space trajectories $\mathbf{K}$ to the feasible set $\mathcal{Q}_{\alpha,\beta}^{N_c}$. This results in a projected gradient descent based optimization of the loss function $\mathcal{L}$, which is given by the following updating step for the k-space trajectories $\mathbf{K}$:

$$\mathbf{K}^{t+1} = \Pi_{\mathcal{Q}_{\alpha,\beta}^{N_c}} \left(\mathbf{K}^t - \eta^t \nabla_{\mathbf{K}} \mathcal{L}_r(\mathbf{x}, \widehat{\mathbf{x}})\right) . \tag{7.10}$$

The projected gradient descent formulation gives an equivalent result to optimizing the original reconstruction error $\mathcal{L}_r$, with indicator function of the feasible set $\mathcal{Q}_{\alpha,\beta}^{N_c}$ as the penalty term. However, as the indicator function is non-differentiable, direct use of such a penalty term in auto-differentiation frameworks (as an alternative to projection step as shown in Eq. (7.10)) generates sub-gradients which makes the optimization process extremely slow, due to oscillations as there are multiple sub-gradients at each evaluation point.

Practical implementations involved 50 iterations of the projection algorithm from [Cha+16] which was sped up using GPU implementations as shown in [Cha+22d]. In practice, benchmarking with a very small reconstruction network (NC-PDNet with 3 iterations, rather than 12) showed 2.25 seconds per step for penalty-based schemes, while with projection, the computation time was 3.16 seconds per step.

7.2.8 Practical implementations

All our implementations in 2D were carried out on V100 GPU with 32GB memory, while our 3D implementations needed the next generation A100 GPUs with 80 GB of memory. Most of the memory in 3D was occupied by the activations from the 3D convolutional neural networks used in the image denoising step in NC-PDNet. Memory efficient implementations of NUFFT was carried out by using *tensorflow-nufft* [Mon22], which is based on tensorflow implementations of *cuFINUFFT* [Shi+21].

7.3 Results

In this section, we first compare our results with state-of-the-art methods, particularly BJORK [Wan+21] and PILOT [Wei+20]. Next, we provide an explanation on why our approach outperforms its competitors. In short, the reason is tightly linked to the use of a projection step in the optimization process for enforcing the hardware constraints rather than using penalty terms in the loss function. Finally, we benchmark our jointly learned k-space sampling pattern and reconstruction network in 3D by compare it to SPARKLING trajectories with a learned neural network for image reconstruction.

7.3.1 Comparison with state-of-the-art methods in 2D

We learned k-space trajectories with $N_c = 16$ shots and $N_s = 512$ samples per shot (observation time $T_{\text{obs}} = 5.12$ms, raster time $\Delta t = 10$µs, dwell time $\delta t = 2$µs). For comparison with an earlier baseline, we used SPARKLING trajectories generated with the learned sampling density using LOUPE [Bah+20a] as obtained in [CRC21] and trained NC-PDNet [Ram+22b] as a reconstruction model for it.

We compared our results with PILOT and BJORK trajectories, which were obtained directly from the respective authors. As we didn't receive their trained reconstruction networks, we trained an NC-PDNet by ourselves for a fair comparison: NC-PDNet makes use of DC and its Cartesian version stood 2nd in the 2020 fastMRI challenge [Muc+21a]. This way, we used the same reconstruction neural network for all the trajectories (with the same parameters), which was trained individually. Our comparison with PILOT (Fig. 7.3) was carried out for T_1 and T_2 weighting contrasts in the fastMRI data set.

As the BJORK trajectory was learned for $\Delta t = 4$µs, to ensure fair comparison, we obtained trajectories with the same specifications. This comparison (Fig. 7.4) was done at different Undersampling Factor (UF) $= \frac{N^D}{N_c \times N_s}$. Note that UF is a measure of how much the k-space is under-sampled with respect to fully sampled Cartesian k-space, while AF reflects on how fast the scan is with respect to the Cartesian reference scan.

We first proceed to analyze the k-space trajectories as compared to those yielded by BJORK and PILOT. Then, we compare the reconstruction results of the learned trajectories with BJORK and PILOT.

Trajectory analysis

When looking at the zoomed portions of optimized trajectories in Fig. 7.2, we observe that PILOT has a hole at the center of k-space (cf. the white spot shown in the bottom inset) while BJORK samples the k-space densely slightly off the center (cf. bottom inset), which is suboptimal. In contrast, PROJeCTOR and SPARKLING methods sample the central region of k-space more densely, which could help obtain improved image quality, notably the contrast.

We also observe at the bottom of each panel in Fig. 7.2 that PILOT and BJORK do not use the hardware gradient capacities at their maximum values and have similar gradient ($\mathbf{G}(t)$) and slew rate ($\mathbf{S}(t)$) profiles, while SPARKLING and PROJeCTOR trajectories, are hitting the gradient constraints more often for the maximal gradient and almost everywhere for the slew rate. This difference could be attributed to using a projector for handling hardware constraints in PROJeCTOR and SPARKLING as compared to handling them through penalty terms in PILOT and BJORK.

Retrospective study

Next, we compared the results of image reconstruction from retrospectively under-sampled k-space data using PILOT (Fig. 7.3) and BJORK (Fig. 7.4) trajectories. To this end, we used 512 slices from fastMRI validation data set. We observe that both SPARKLING with a learned density and PROJeCTOR outperform PILOT and BJORK, with PROJeCTOR yielding the best scores with a gain of nearly 0.06 in SSIM and 3-4dB in PSNR values as compared to PILOT and BJORK. We computed paired t-tests on Structural Similarity Index Measure (SSIM)/Peak Signal-to-Noise Ratio (PSNR) scores between PILOT and PROJeCTOR on one hand and BJORK and PROJeCTOR on the other hand and obtained p-values $p < 10^{-4}$, thus confirming that the improvements we observed visually and quantitatively are statistically significant.

7.3.2 Hardware Constraints: Penalty vs Projection

In the above section we showed how our method outperforms PILOT and BJORK in terms of reconstructed image quality. We assume that these results are due to the different manner the hardware constraints on the gradients are enforced in the learning process (projector vs regularizer). To validate this hypothesis, we learned 3D hardware compliant k-space sampling trajectories through joint optimization with a reconstruction network using a penalty term instead of a projector.

In Fig. 7.5, we present the learned hardware compliant k-space sampling trajectories using the projection and penalty-based methods and then in Fig. 7.6 we depict their corresponding slew rate and gradient profiles. Additionally, we also show in Fig. 7.6 the validation SSIM scores as a function of the penalty weight (λ). For the sake of simplicity, we assume $\lambda = \lambda_i$, $i \in \{1, 2, 3\}$ and we obtain results for $\lambda = 10^{-3}$, which is the lowest level of penalty resulting in hardware-compliant trajectories at the end of training. By doing so, we ensure that we do not influence too much the trajectory shape. However, in our grid search experiments of varying λ across different orders of magnitude, we did not observe any significant drop in validation loss within the range $[10^2, 10^{-3}]$. Further, to obtain an insightful baseline, we also obtain results for $\lambda = 0$ corresponding to non-admissible trajectories as we do not enforce any penalty on the gradients and slew rates. Last, we also display the learned trajectories using the PROJeCTOR.

We observed that the best reconstructed image quality can be obtained for $\lambda = 0$ in terms of validation SSIM and PSNR scores. Further, increasing the weight λ of penalty terms, the validation SSIM and PSNR scores drop as the k-space trajectories get more constrained. Interestingly, as $\lambda = 10^{-3}$ the k-space trajectories are getting hardware compliant (see Fig. 7.6(B)(iii))), but they become strongly constrained and do not reach the same level of flexibility as those learned by PROJeCTOR. This results in a significant decrease in the performance of penalty-based method as compared to projection-based methods.

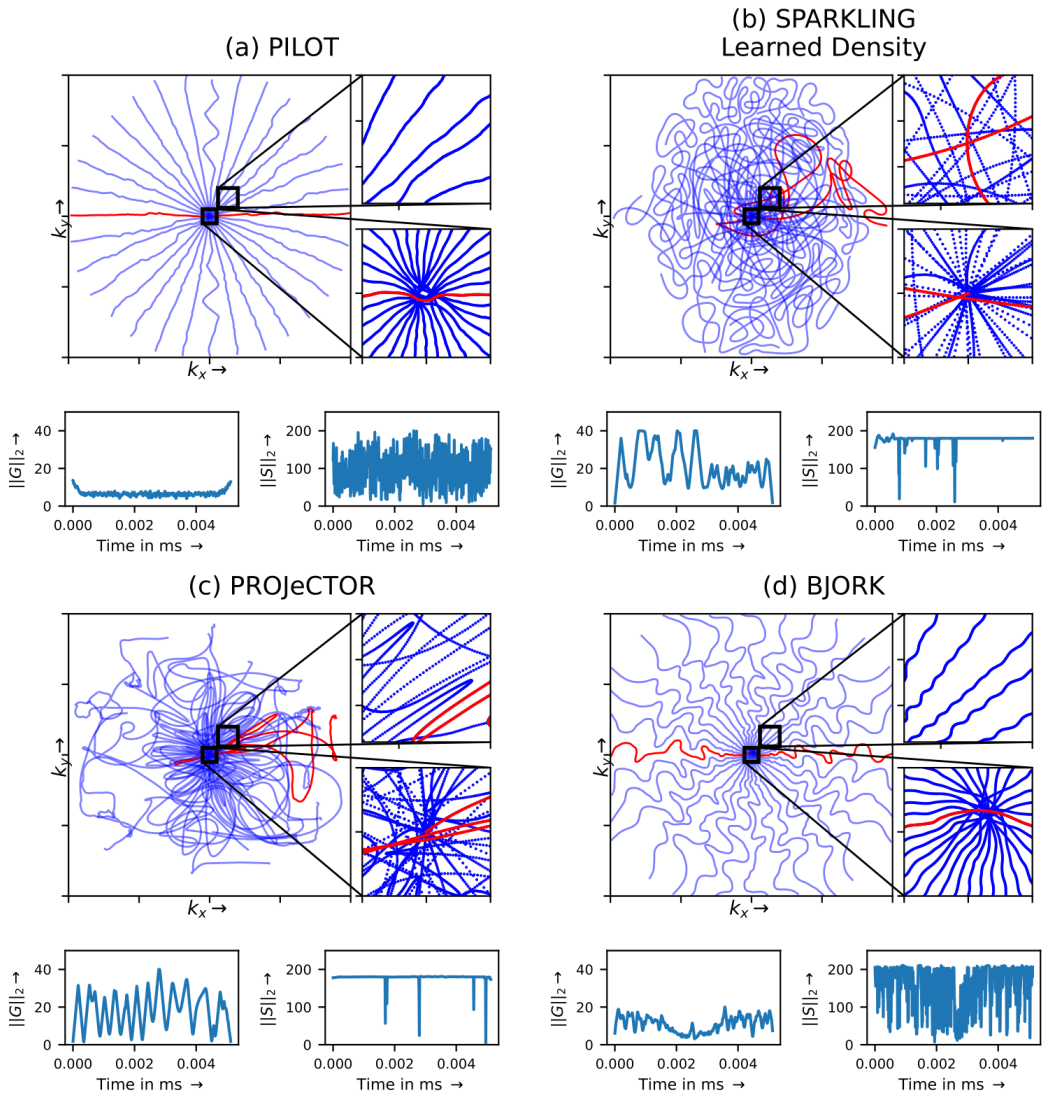


Figure 7.2 – The optimized hardware compliant non-Cartesian k-space trajectories using **(a)** PILOT, **(b)** SPARKLING with learned density using LOUPE, **(c)** PROJeCTOR scheme, **(d)** BJORK. The number of shots is N_c=16. The number of dwell time samples are set to match the same number of sampling points overall. Zoomed in visualizations of the center of k-space (bottom) and slightly off-center (top) is presented at the right of corresponding trajectories. The ℓ_2 norm of the corresponding gradient $||G||_2$ (in mT/m) and slew rate $||S||_2$ (in T/m/s) profiles are depicted below each trajectory.

Finally, we observed that using projection-based method, the k-space trajectories are closer those obtained with $\lambda = 0$.

7.3.3 Comparison with SPARKLING in 3D

Finally, we compared the performances of our data-driven jointly learned k-space trajectories to the model-driven SPARKLING trajectories in 3D imaging. The networks were trained for 240 epochs, with 32 steps per epoch on the Calgary brain data set [Sou+18], for trajectories at AF=20, resulting in $N_c = 1681$ shots. To ensure a fair comparison, we learned the same NC-PDNet, i.e. image reconstruction neural network for the same number of steps as was done for PROJeCTOR trajectories.

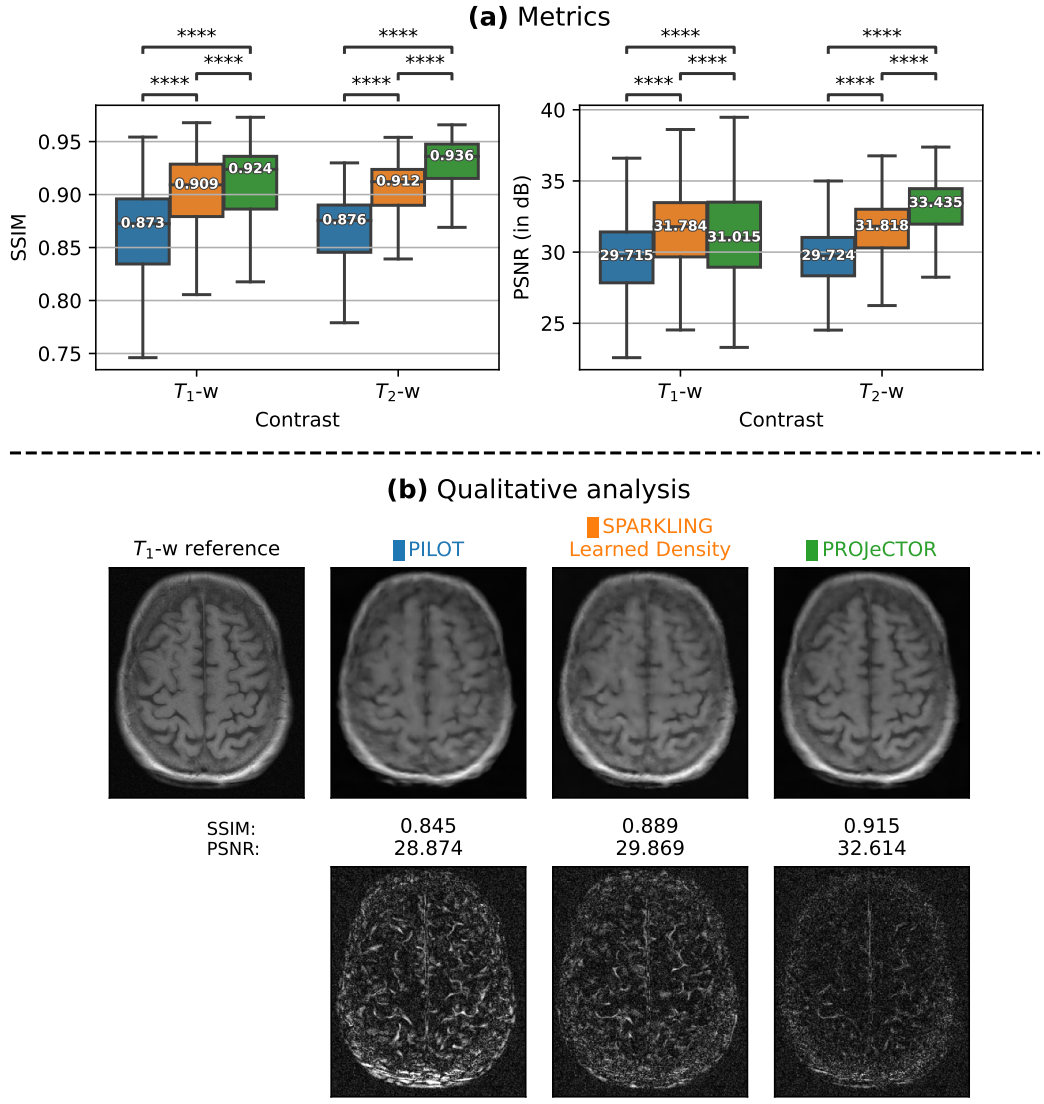


Figure 7.3 – **(a)** Box plots comparing the image reconstruction results on a retrospective study at UF=2.5 ($N_c = 16$, $N_s = 512$, $\frac{\Delta t}{\delta t} = 5$) using 512 slices of T_1 and T_2 contrasts (fastMRI validation data set) using PILOT (blue), SPARKLING with learned density (orange) and PROJeCTOR (green). SSIMs/PSNRs appear at left/right. The median values of these metrics are highlighted inside the box plots. The significance levels are indicated as paired t-test and are all significant with $p < 10^{-4}$. **(b) Top:** T_1-w reference image and reconstruction results for a single slice from `file_brain_AXT1PRE_209_6001221.h5` with corresponding strategies. **(b) Bottom:** The residuals maps, scaled to match and being comparable across methods.

From the mid-slice cuts of gridded sampling patterns in k-space in Fig. 7.7[A-B](b)-(d), we see that SPARKLING trajectories present radial-like sampling at the center of k-space which could induce some k-space holes (see red arrows in (A.b) and (A.c)). This type of imperfections is not present in the learned PROJeCTOR k-space sampling pattern ((B.b) and (B.c)). Further, as the trajectories and reconstruction network were learned on partial Fourier k-space data, PROJeCTOR trajectories also learned to exploit this by not sampling these regions (see the dark areas pointed by green arrow in (B.b) and (B.d)).

Finally, comparing the actual reconstructed MR images in Fig. 7.8, we see that SPARKLING trajectories result in blurrier images, while PROJeCTOR retains the high-frequency

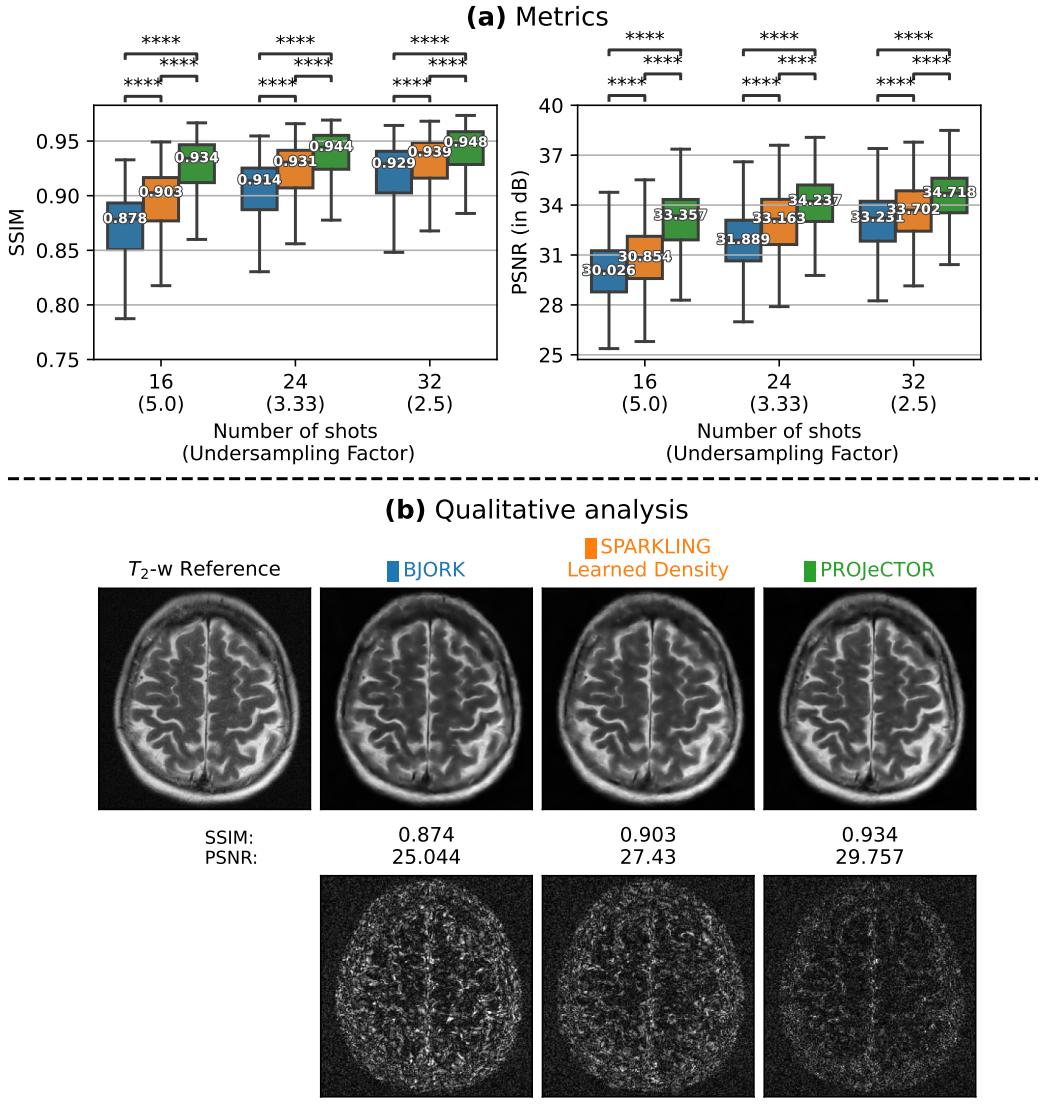


Figure 7.4 – **(a)** Box plots comparing the image reconstruction results on a retrospective study using 512 slices on T_2 contrast (fastMRI validation dataset) using BJORK (blue), SPARKLING with learned density (orange) and PROJeCTOR (green). The median values of these metrics are highlighted inside the box plots. We present the results at varying UF characterized with $N_c = 16$, 24 and 32. SSIMs/PSNRs appear at left/right. The significance levels are indicated as paired t-test and are all significant with $p < 10^{-4}$. **(b) Top:** T_2-w reference image and reconstruction results for a single slice from `file_brain_AXT2_205_2050175.h5` with corresponding strategies. **(b) Bottom:** The residuals maps, scaled to match and being compare across methods.

details. This can be observed qualitatively through the residual images and quantitatively through box plots indicating SSIM and PSNR scores, taken on 20 test data sets. We see that PROJeCTOR outperforms SPARKLING by nearly 0.02 points in SSIM and +2dB in PSNR scores. As our evaluation is done on 20 matched data points, we use Wilcoxon signed-rank test, which is a non-parametric statistical hypothesis test used here to compare the locations of two populations using two matched samples. We found that the difference in both the SSIM and PSNR scores are statistically significant with $p < 10^{-5}$.

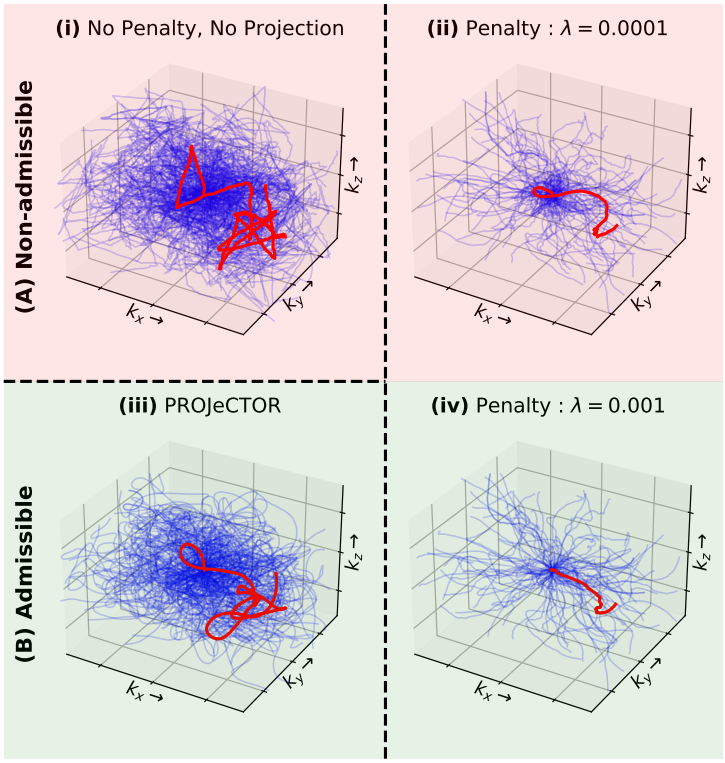


Figure 7.5 – Comparison of **(iii)** PROJeCTOR trajectories with respect to penalty-based versions (**(ii)** and **(iv)**). The **(A)** non-admissible trajectories are shaded in red while admissible trajectories are shaded in green. Also, unconstrained (no penalty and no projection) trajectories are also presented in **(i)**.

7.4 Discussion

In this work, we present a generic framework for jointly learning the trajectory and image reconstruction neural network. We embedded the projection step from [Cha+16] and learned these PROJeCTOR trajectories through a novel projected gradient descent fashion to ensure hardware compliance.

Although the learned neural networks in PILOT [Wei+20] and BJORK [Wan+21] were not available for a full end-to-end comparison, we performed a fair assessment by training a NC-PDNet [Ram+22b] as a common deep neural network reference for image reconstruction. Through retrospective studies in 2D on the fastMRI validation data set, we showed that PROJeCTOR works across multiple resolutions and leads to superior performance of the trajectories and improved image quality overall, with nearly 3-4dB gain in PSNR value and almost 0.06 gain in SSIM score.

This improvement over state-of-the-art methods can be attributed to the embedded projection step as compared to penalty to ensure hardware compliance. We carried out an ablation study and showed that the projection step is crucial for having significantly improved performance of the learned trajectories, as compared to penalty-based approaches.

Finally, in 3D we compared the model-driven method SPARKLING with the data-driven method PROJeCTOR and showed a gain of 2dB in PSNR and 0.02 gain in SSIM in favor of the latter.

Future prospects of this work include prospective implementations through modifications of T_1 and T_2-w imaging sequences. Such practical implementations could possibly bring up new sequence-specific constraints on k-space trajectories and also affect the overall performance due to lower Signal-to-Noise Ratio (SNR).

A limitation of current work is that our training paradigm was setup in emulated single coil

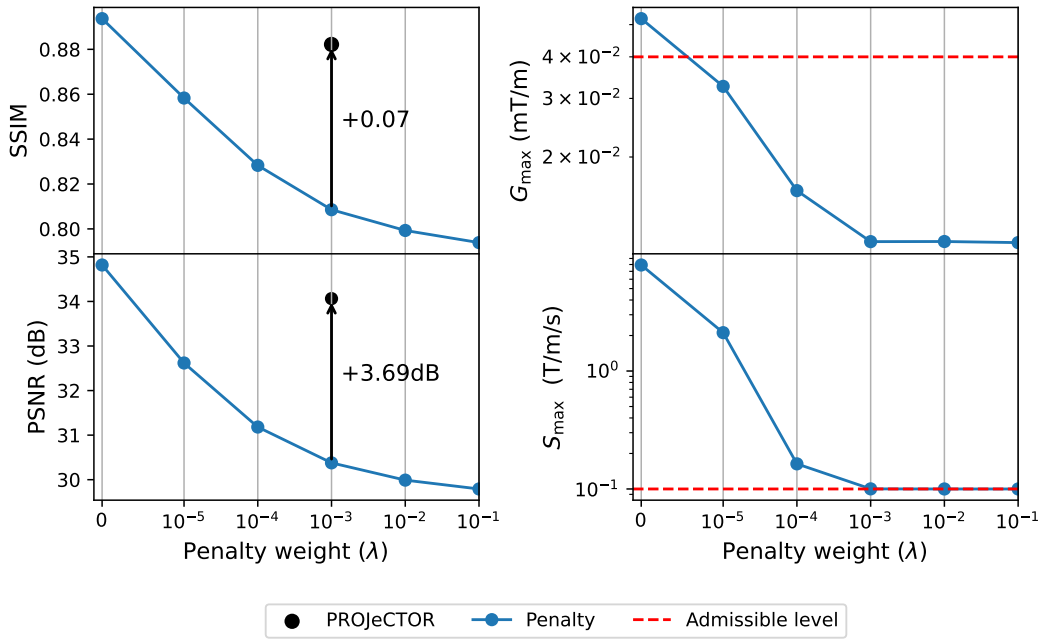


Figure 7.6 – **Left:** Performance metrics in SSIM and PSNR of penalty-based method at varying penalty weight λ. The performance PROJeCTOR is also shown for comparison. **Right:** The feasibility of the penalty-based learned k-space trajectories at varying penalty weights (i.e. hyper-parameter) λ, shown by maximum slew rate $S_{\max}$ and maximum gradient strength $G_{\max}$ in the entire sampling pattern. The respective admissible upper levels are drawn with a red dotted line.

setting as we were limited by memory constraints on GPU. A more realistic implementation would involve multi-coil imaging setting is mandatory to efficiently utilize parallel imaging and get closer to the real data acquisition context, allowing us to reach higher AF. However, this memory bottleneck can be alleviated through efficient transfers between CPU and GPU or multi-GPU implementations. Further, the network can be improved by extending the currently implemented simple forward acquisition model NUFFT to a more realistic and complex model which takes off-resonance effects due to B_0 inhomogeneities [Dav+22] and gradient imperfections into account. These aspects will be explored in our future works.

* * *
* *
*

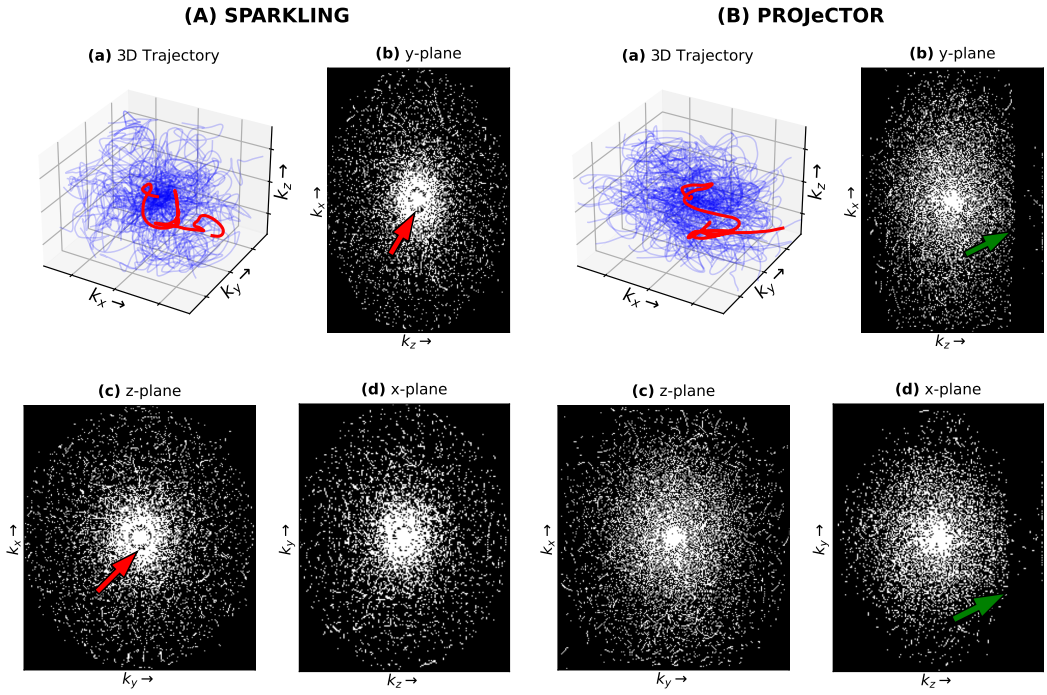


Figure 7.7 – k-space sampling trajectories for **(A)** SPARKLING and **(B)** PROJeCTOR. For easier visualization, only 70 shots of 3D trajectory are shown in **(a)**. The resulting gridded sampling pattern is shown for mid-plane slices along **(b)** y-plane, **(c)** z-plane and **(d)** x-plane.

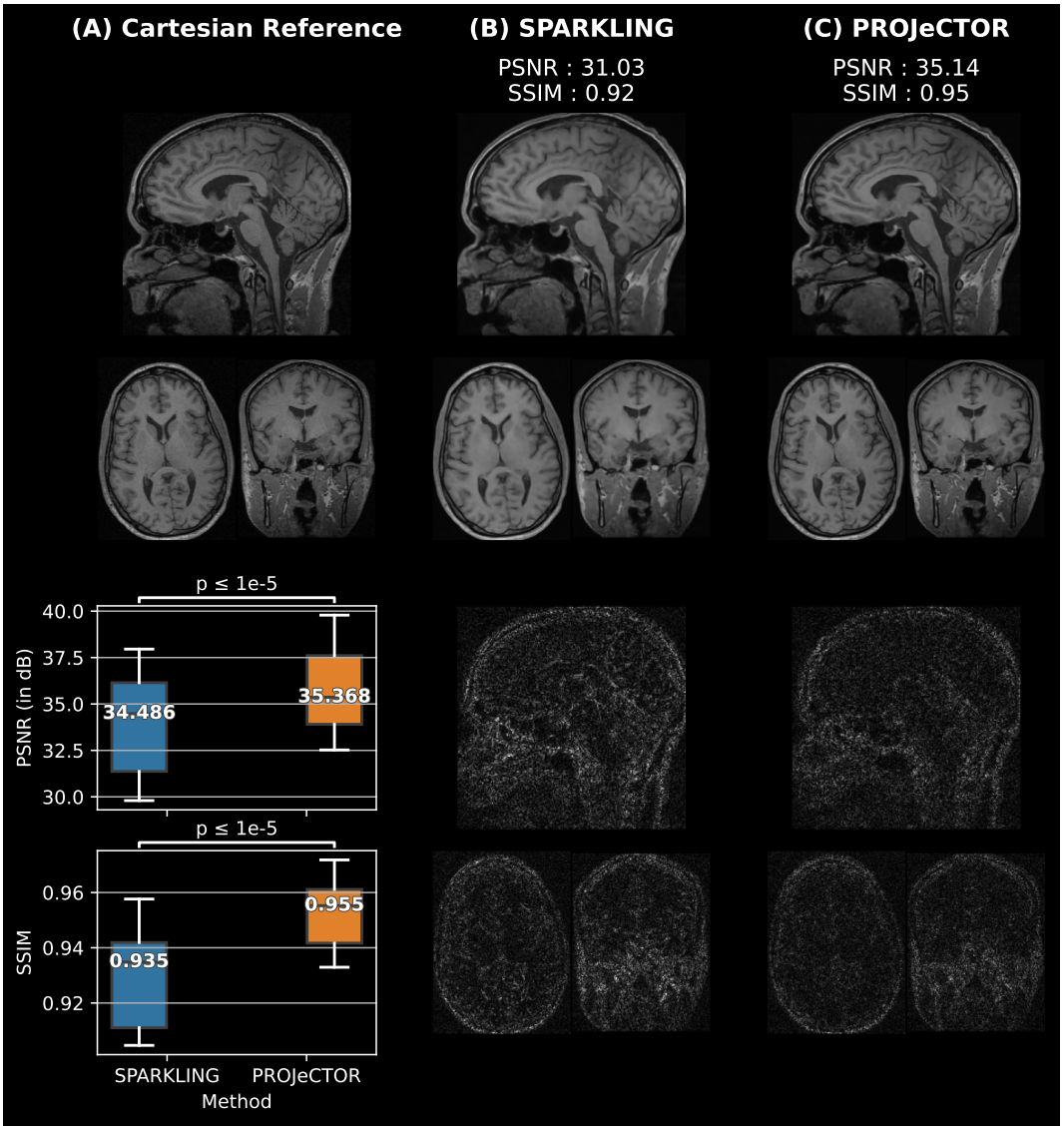


Figure 7.8 – Qualitative and quantitative comparisons of reconstructed images from 3D **(B)** SPARKLING and **(C)** PROJeCTOR trajectories at AF=20 as compared to **(A)** Cartesian reference. The reconstructed images are shown in top row, while the residuals are shown in the bottom. Further, box plots of SSIM and PSNR scores on 20 test data sets are shown in the bottom-left. The significance levels are marked through paired samples Wilcoxon test.

Chapter 8

Conclusions and Perspectives

DURING the course of this thesis, we significantly contributed to the Magnetic Resonance Imaging (MRI) community, particularly in the design of hardware compliant non-Cartesian k-space sampling trajectories that yield improved reconstructed image quality. These newly developed trajectories were further extending to yield improved images on prospectively acquired data through Minimized Off Resonance Effect (MORE) and Gridding of Low Frequencies (GoLF). Finally, in the last stage of the thesis, with a paradigm shift from model-based design to data-driven learning of trajectories jointly with reconstructor, we demonstrated we can speed up MRI scans by an order of 15-20x as compared to Cartesian reference, without much degradation in the reconstructed image quality.

8.1 Contributions and limitations

At the beginning of the thesis, the Spreading Projection Algorithm for Rapid K-space sampLING (SPARKLING) algorithm developed in-house was sped up through use of Fast Multipole Methods (FMM) and parallel implementations on multicore CPUs and GPUs. With this, the SPARKLING algorithm was fully extended to optimize 3D hardware compliant k-space sampling trajectories, which efficiently under-sampled the k-space in all the three dimensions, a significant challenge prior to my arrival in the team due to the underlying computational bottleneck. The optimized trajectories yielded peaky Point Spread Function (PSF) which was characterized and quantified using PSF metrics: lower Full Width at Half Maximum (FWHM), and higher Peak-to-Sidelobe Level (PSL) and Peak-to-Noise Level (PNL). Retrospective studies were carried out to grid-search optimal Target Sampling Density (TSD), which was an input to the SPARKLING algorithm. Finally, these trajectories were applied for the case of prospective 3D Susceptibility Weighted Imaging (SWI), which allowed 2.5-3.75x shorter scan times compared to GRAPPA-4 parallel imaging acquisition at 3 Tesla without compromising image quality. However, in our prospective experiments, we observed

strong off-resonance artifacts in the reconstructed images, which were not observed in the simulation based retrospective experiments. While physics informed post-processing based techniques were developed in parallel in the team [Dav+22] in order to reduce the effects of off-resonance artifacts without needing additional scans for acquiring the ΔB_0 field map, these techniques were computationally expensive and hence not really viable for clinical applications.

We noticed that these strong off-resonance effects observed with SPARKLING trajectories were particularly due to arbitrary readout paths used by optimized trajectories in the k-space. Such arbitrary trajectories lead to temporally discontinuous sampling of the k-space which ignore the temporal nature of the MRI signal being sampled, thereby amplifying the off-resonance artifacts in the reconstructed images. To overcome this issue, MORE-SPARKLING was introduced through a new repulsion term that temporally constrains the k-space samples and results in a temporally smooth under-sampling of k-space. In our studies with both phantom with artificially degraded shim and *in vivo* acquisitions, these new trajectories recovered the signal losses at locations of strong off-resonance effects. Another concern with SPARKLING algorithm involved the Echo Time (TE) point constraint which resulted in oversampling the center of k-space much beyond the Nyquist limit. This suboptimality was tackled through GoLF, which involved additional affine constraints to yield trajectories which fully sample the center of k-space optimally using Cartesian sampling, and have good coverage in periphery through non-Cartesian sampling. Through this compound sampling approach, the reconstructed images were significantly improved as compared to vanilla SPARKLING trajectories, which allowed for further reducing the scan time by 5x as compared to the same GRAPPA-4 acquisition at 3 Tesla.

One major limitation of the SPARKLING algorithm was the need for TSD as input, which was obtained through grid-searching over a range of values for a heuristically parameterized density inspired by Compressed Sensing (CS) literature. With the rise of MRI datasets with raw k-space data like fastMRI [Zbo+18], we shifted our focus from such a model-based designed to data-driven learning of the TSD. To this end, we setup four different candidates for optimized TSD and design SPARKLING trajectories accordingly. Reconstruction was performed using NC-PDNet [Ram+22b], which was separately trained with the SPARKLING trajectories with these candidate densities. Through our retrospective results performed on the validation fastMRI dataset, we observed that the Learning-based Optimization of the Under-sampling PattErn (LOUPE) [Bah+20b] based sampling density yielded the best performance in terms of image quality metrics.

The gridded LOUPE based sampling density was learned by training in the Cartesian domain, and then used as input to generate non-Cartesian SPARKLING trajectories. Later for validation purposes, these trajectories were used again to learn a non-Cartesian reconstruction network. This two-step process is computationally expensive and also suboptimal as the training paradigm was different as compared to the validation setting. To address this lack of consistency, we proposed to directly learn non-Cartesian k-space sampling trajectories. To this end, we developed a generic framework for jointly learning the trajectory and image reconstruction neural network, while embedding the projection step from [Cha+16] to ensure hardware compliance of the learned trajectories. These newly learned PROjection for Jointly lEarning non-Cartesian Trajectories while Optimizing Reconstructor (PROJeCTOR) trajectories yielded improved retrospective performance as compared to earlier state-of-the-art methods Physics-informed learned optimal trajectories (PILOT) and B-spline parameterized Joint Optimization of Reconstruction and K-space trajectories (BJORK), with nearly 3-4dB gain in PSNR value and almost 0.06 gain in SSIM score. Through an ablation study, we showed that such improvement can be attributed to projection step as compared to penalty terms that are used in concurrent approaches to ensure hardware compliance. Finally, in 3D we compared the model-driven method SPARKLING with the data-driven method PROJeCTOR and showed a gain of 2dB in Peak Signal-to-Noise Ratio (PSNR) and 0.02 gain in Structural Similarity Index Measure (SSIM) in favor of the latter. However, this method was purely limited to retrospective studies and the learned network was primarily trained on emulated single coil data, resulting in trajectories which do not exploit parallel imaging strategies during acquisition.

8.2 Perspectives

Several perspective directions can be explored to further improve the methods that were developed in the context of this thesis. Although this work shifted from model-based SPARKLING to data-driven PROJeCTOR, we firmly believe that both are equally important and can to be explored further independently. However, the methods and findings discovered in one frameworke can also be exploited and transposed to improve the other. Some of these directions will be soon explored by my peers in the CS-MRI group at NeuroSpin. We briefly discuss the perspectives for SPARKLING and PROJeCTOR trajectories below:

8.2.1 Model-based SPARKLING

Extension to 3D+t or 4D

The SPARKLING trajectories can be extended to 3D+time, resulting in trajectories that satisfy the TSD for every fixed time window of scanning, while maintaining incoherence across time windows. This can be particularly useful for dynamic MRI like Cardial imaging and Functional MRI (fMRI). Preliminary studies of applying SPARKLING for fMRI with simple scan-and-repeat protocol showed promising results, where results are comparable to the state-of-the-art Echo Planar Imaging (EPI) [Amo+22c; Amo+22b; Amo+22a].

Application to other modalities which need ultra low echo times

As the SPARKLING algorithm can also be initialized with center-out trajectories, they can be used for imaging applications that need ultra short echo times of acquisition. In such applications, MORE-SPARKLING trajectories can result in improved reconstructed image qualities as compared to current state-of-the-art including Twisted Projection Imaging (TPI) and radial kooshball [Lar+08]. Currently, this method is being explored for magnetic resonance spectroscopic imaging (MRSI), particularly for Sodium (Na^{23+}) imaging [Bap+22a].

Compound sampling trajectories

By embedding affine constraints in the SPARKLING framework, we introduced novel compound sampling approach to optimize k-space sampling trajectories having both Cartesian and non-Cartesian parts to extract the best of both worlds. However, its implementation in this work was purely done to fully sample the center of k-space using Cartesian sampling and therefore there is room for improvement through the following extensions:

- **Cartesian + non-Cartesian sampling pattern:** We can extend the above framework to have more complex sampling trajectories which overall result in a Cartesian and non-Cartesian sampling pattern, to reduce the impact of inaccuracies from Nonuniform Fast Fourier Transform (NUFFT) operator and accelerate the reconstruction process.
- **Towards EPI+SPARKLING :** The proposed GoLF-SPARKLING trajectories can be extended by having every trajectory pass through the center of k-space multiple times as different Cartesian lines, resulting in a larger portion of center of k-space sampled with Cartesian sampling. The optimized trajectories in such cases would provide an intermediate behavior between fully non-Cartesian sampling trajectories and EPI, allowing us to better understand the trade-off between the two.
- **Sensitivity maps and wavelet coefficients:** As the center of k-space is fully sampled, we can quickly obtain a low resolution estimate of the reconstructed image through simple Inverse Fast Fourier Transform (IFFT). Such low resolution images can be good candidates for the approximation coefficients in the wavelet domain of the reconstructed image, of course in the context of CS reconstruction. Further, they can be used to extract fast and reliable estimates of sensitivity maps.
- **Acceleration through parallel imaging:** We can now incorporate parallel imaging methods like GRAPPA [Gri+02], SENSE [Pru+99] and CAIPIRINHA [Bre+05] to

further increase the Acceleration Factor (AF), while maintaining or even improving excellent image clarity. sampling

Handling PNS constraints

The developed trajectories in this thesis exploit the gradient hardware of the scanner, leading to high gradient amplitudes played on the scanner rapidly. This results in Peripheral Nerve Stimulation (PNS) issues, which was reported by some volunteers and also sometimes the gradient safety watchdog (GSWD) of the scanner was triggered, leading to the scan being aborted. The SAFE model [HG00] can model the working of this GSWD, which can be used to predict the PNS associated with a given trajectory and hence extend the set of hardware constraints on the trajectories by adding physiological ones. However, as of now no projector exists for these PNS constraints, the difficulty lying in the fact they are not necessarily convex. Although penalty terms can be used to control PNS, as proposed in SNOPY [Wan+22], such penalty terms can however lead to sub-optimal trajectories as shown in Chapter 7. In order to prevent such PNS issues on the scanner, we generated trajectories with a reduced maximum allowable slew rate of 100T/m/s, as compared to 200T/m/s allowed by the scanner hardware. While no PNS issues were reported in all our studies with this lower slew rate, the trajectories are more constrained than required, hence suboptimal. To address this concern, there is a need to develop a PNS projector, which can be used to generate trajectories with higher slew rates, while still maintaining the PNS constraints.

Trajectories for motion correction

The SPARKLING algorithm can be tweaked to have a good navigator within each shot, which can be helpful in *prospectively* detecting, analyzing and correcting motion during the scan.

8.2.2 Data-driven PROJeCTOR

Prospective studies

The currently developed PROJeCTOR trajectories need to be tested in prospective setting to validate their performances in real world scenarios. Any observed gap needs to be accounted for and improved through additional constraints on the trajectory or modeling.

Better forward modelling

The generic framework developed in this work to learn trajectories relies on simple NUFFT operator for forward modeling. However, more realistic forward modeling can be achieved by also incorporating the effects of field inhomogeneities, motion and eddy currents. Further, realistic noise models can be incorporated to better understand the impact of noise on the learned trajectories.

Extension to multi-coil setting

As discussed earlier the current implementation of PROJeCTOR is limited to emulated single coil data due to memory constraints on GPU. A more realistic implementation that would involve a multi-coil imaging setting, is mandatory to efficiently utilize parallel imaging and get closer to the real data acquisition context, allowing us to reach higher AF. However, this memory bottleneck can be alleviated through efficient transfers between CPU and GPU or multi-GPU implementations. This perspective remains a significant step forward.

Inclusion of MORE and GoLF features

As observed with SPARKLING trajectories, the addition of MORE and GoLF features can significantly improve their performances. These features are generic and can also be

incorporated into the PROJeCTOR trajectories, allowing us to anticipate an improved image quality in this extended learning framework.

* * *
* *
*

Appendices

Appendix A

3D SPARKLING

A.1 Trajectory

A.1.1 Gradients and Slew Rates

We present the gradients and slew rates obtained for a single shot of full 3D Spreading Projection Algorithm for Rapid K-space sampLING (SPARKLING) trajectory in Fig. A-1. We show that the trajectory was mostly slew rate constrained, thereby making the percentage of readouts with gradient magnitude constraint active to be close to 0. The gradient waveform was never saturated but achieved its maximum nearby the center of k-space as shown in Fig. A-1(a). This is because all the trajectories pass through the center of k-space, thereby drastically increasing the sampling density of the sampling pattern. Due to this, each trajectory moves at highest velocity to achieve lesser k-space sample points here, thereby achieving the Target Sampling Density (TSD). Note that although the scanner slew rate constraint is 200 T/m/s, the trajectories were optimized with a tighter constraint of maximum allowed slew rate of 180 T/m/s, which explains the lower slew rate limits observed in the trajectory in Fig. A-1(b).

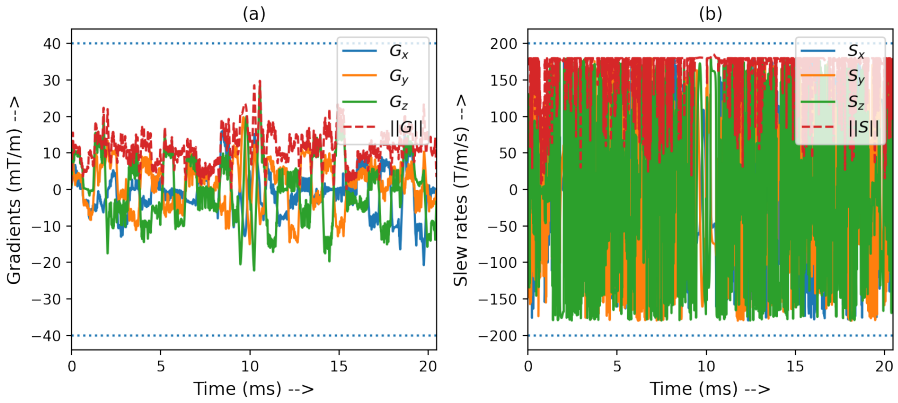


Figure A-1 – The **(a)** gradients and **(b)** slew rates for a single shot from a Full 3D SPARKLING trajectory with AF=20. We have also marked the Scanner hardware constraints ($G_{\max}$ = 40mT/m and $S_{\max}$ =200 T/m/s) with black dotted lines.

A.1.2 Off-resonance and T_2^* decay

We carried out off-resonance and T_2^*-decay simulations on the trajectory and analyzed the Point Spread Function (PSF) in Fig. A-2. We systematically added a T_2^*-decay by taking a constant value of T_2^* = 30ms, and also performed off-resonance simulations by adding a constant off-resonance ($\Delta B0$) of 25Hz. In Fig. A-2 we observe that the noise level in the

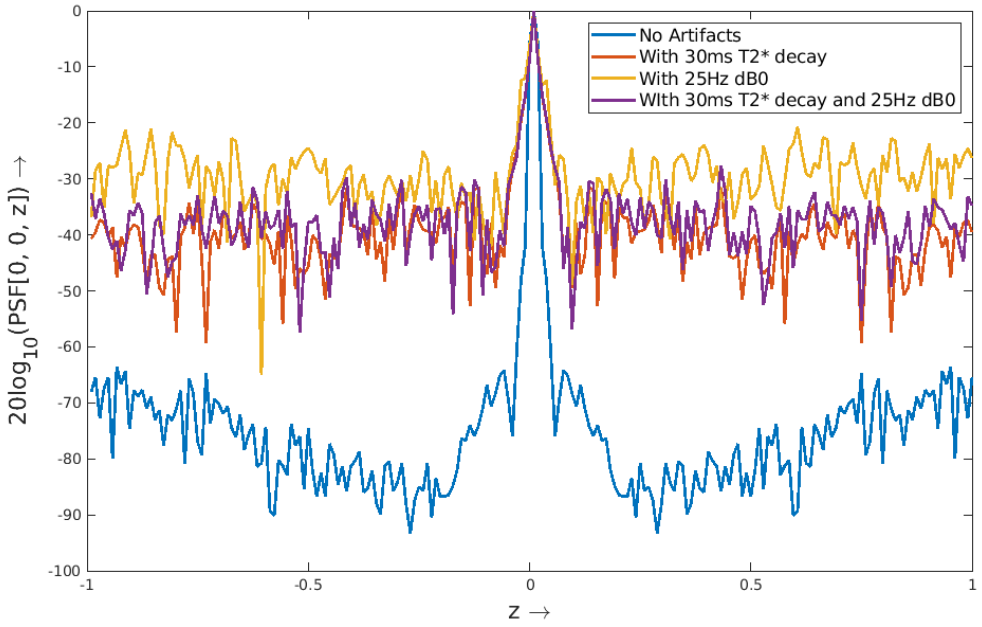


Figure A-2 – Simulated effects of T_2* decay (of 30ms) and constant $\Delta B0$ (of 25Hz) on the point spread function of the AF=10 full 3D SPARKLING trajectory. All the PSFs are normalized such that the maximum value is 1, for easier visual comparison.

PSF is significantly increased (by $\simeq 25-30$ dB) when adding $\Delta B0$ and T_2^* decay. Further, we see an increase in sidelobe level when adding $\Delta B0$, which however is reduced when combined with T_2^* decay. We observe that the effective PSF is spread under T_2^* and Bo inhomogeneities, leading to drop in effective resolution, which is studied in depth in the core paper in Sec. 4.4.4. This study reveals some of the reasons that explain the gap in image quality between retrospective and prospective results, especially for in vivo acquisitions, where we have spatially varying T_2^* and $\Delta B0$ (i.e. $T_2^*(\mathbf{r})$ and $\Delta B0(\mathbf{r})$ maps).

A.1.3 Eddy Current and Trajectory Measurement

As the presented trajectory rapidly explores the k-space, it is vital to ensure that the MRI scanner is able to play the complicated gradient waveforms in Fig. A-1 with minimal errors. These errors could be induced by eddy currents and gradient imperfections. To this end, we ran the AF=20 full 3D SPARKLING trajectory on an Investigative 7T MR System (MAGNETOM 7T, Siemens Healthcare, Erlangen, Germany) and measured the trajectory with the SKOPE dynamic field camera [De +08]. We used a 7T scanner for this study due to compatibility issues at 3T. However, we do not expect drastic changes in our results as the gradient system is the same for both scanners. We present the theoretical (i.e. prescribed by the 3D SPARKLING algorithm) and measured trajectories for 3 random k-space shots in Fig. A-3. Further, we quantitatively measured the error as to be 0.0016 ± 0.0012 (with the k-space normalized to $\Omega \in [-1, 1]^D$).

A.2 MR image reconstruction

The MR image reconstruction of 3D multi-channel data acquired from phased array receiver coils was carried out using a self-calibrating Compressed Sensing (CS) reconstruction algorithm [El +18a] in the synthesis formulation by solving for the wavelet coefficients $\mathbf{z}$ in (A.1):

$$\widehat{\mathbf{z}} = \underset{\mathbf{z}\in\mathbb{C}^{N_x\times N_y\times N_z}}{\operatorname{argmin}} \frac{1}{2}\sum_{\ell=1}^{L} \|F_\Omega \mathbf{S}_\ell \mathbf{\Psi}^* \mathbf{z} - \mathbf{y}_\ell\|_2^2 + \lambda\|\mathbf{z}\|_1 \tag{A.1}$$

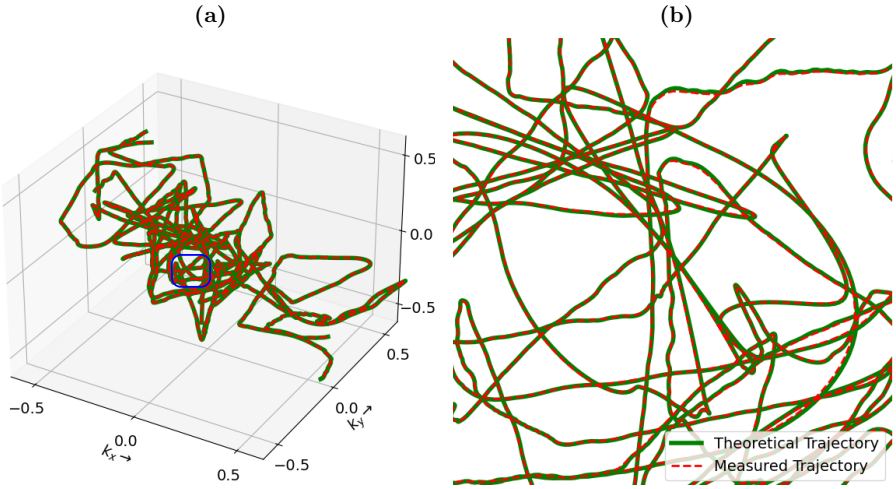


Figure A-3 – **(a)** Three random shots from theoretical and measured trajectories for AF=20 full 3D SPARKLING trajectory. **(b)** Zoomed in at the center of k-space.

where the number of channels was $L = 44$ and $N_x = N_y = 384$ and $N_z = 208$. Here the data fidelity is enforced with SENSE operators $(F_\Omega \mathbf{S}_\ell)_{\ell=1}^{L}$, where F_Ω is the Nonuniform Fast Fourier Transform (NUFFT) operator and $\mathbf{S}_\ell$ is sensitivity map for ℓ^{th} channel estimated by density compensated adjoint of the 20 percent of acquired k-space center (see details in [El +18a]). $\lambda > 0$ is the regularization parameter for promoting sparsity using ℓ_1-norm regularization in the wavelet domain $\mathbf{\Psi}$. For our reconstructions, we used Symlet 8 wavelet with 4 scales of decomposition for $\mathbf{\Psi}$. The regularization parameter λ was grid searched between $(10^{-10}, 10^{0})$ while maximizing for the reconstruction quality using SSIM score in retrospective reconstruction. As the sampling operator was 3D non-Cartesian, the reconstruction problem was severely ill-posed with the forward operator $F_\Omega \mathbf{S_l} \mathbf{\Psi}^*$ having a large condition number, thereby impacting the convergence speed. In order to accelerate convergence, we preconditioned the k-space using density compensation. This translates to adding a preconditioner D in the classical proximal gradient descent algorithm (here we used Faster ISTA (FISTA)):

$$\mathbf{z}^{(k+1)} = \text{soft}_{\lambda\tau}\left(\mathbf{z}^{(k)} - \tau \sum_{\ell=1}^{L} \mathbf{\Psi}\mathbf{S}_\ell^* F_\Omega^H D\left(F_\Omega \mathbf{S}_\ell \mathbf{\Psi}^* \mathbf{z}^{(k)} - \mathbf{y}_\ell\right)\right)$$

where $\text{soft}_{\lambda\tau}$ is the soft threshold operator and τ is the step size. The density compensators D were obtained by 10 iterations of method described in [PM99]. The final MR image is given by $\widehat{\mathbf{x}} = \mathbf{\Psi}^* \widehat{\mathbf{z}} \in \mathbb{C}^{N_x \times N_y \times N_z}$ as $\mathbf{\Psi}$ is a basis.

As the raw data was large (for AF=20, $p = 8,388,608$ k-space points), we needed to utilize memory efficient methods to carry out the SENSE operation. For this, we implemented python wrappers for gpuNUFFT [Kno+14] which implements the NUFFT operator in CUDA and utilizes cuBLAS and cuFFT libraries to be efficient in speed and memory. The implementation of the reconstruction was completely done using `pysap-mri`[1] [Gue+20], the plugin of PySAP [Far+20a] dedicated to MR image reconstruction. Despite being a 3D reconstruction problem, the computation time was just 15-30 minutes on a machine with the same hardware specifications as described earlier in Sec. 4.3.

[1] https://github.com/CEA-COSMIC/pysap-mri

A.3 Target sampling density

In order to choose optimal TSD, we performed a grid search of the cutoff C and decay D parameters in (B.1) (Fig. 4.3). This density search was done through retrospective studies on 100 T_2* complex brain images in 2D to reduce computational complexity as the parameterized density is radially symmetric and can be directly extended to 3D. We present the results of this density search in Fig. A-4 for both single and multi-channel settings through retrospective grid search on corresponding emulated single-channel (obtained by virtual coil combination [Par+14] of multi-channel data) and multi-channel in vivo brain complex-valued data, respectively. In the single coil setting we observe that while ($C=30$, $D=3$) seems optimal, we choose ($C=25$, $D=2$) as it is robust to small changes in density and is on the plateau of the value function. In the multi-channel coil scenario, we observe that ($C=1$, $D=1$) is both optimal and robust. The optimal decay of 1 can be explained by the fact that the k-space information is more spread out in the multi-channel setting thereby allowing sampling density to decay at a lower rate to get a better k-space coverage.

In all our studies, we chose ($C=25$, $D=2$) as target density, as the trajectories are more generic and can be adapted to any coil configuration (as compared to the multi-channel optimal density, which was only adapted to our coil configuration). Also, notice that ($C=25$, $D=2$) is still a reliable tuning in the multi-channel setting.

Additionally, it is worth mentioning that improved image quality can be obtained by further exploring the target density parameterization. Finally, this density can be learned in a data driven manner, as illustrated for T_1 and T_2-w imaging in [CRC21].

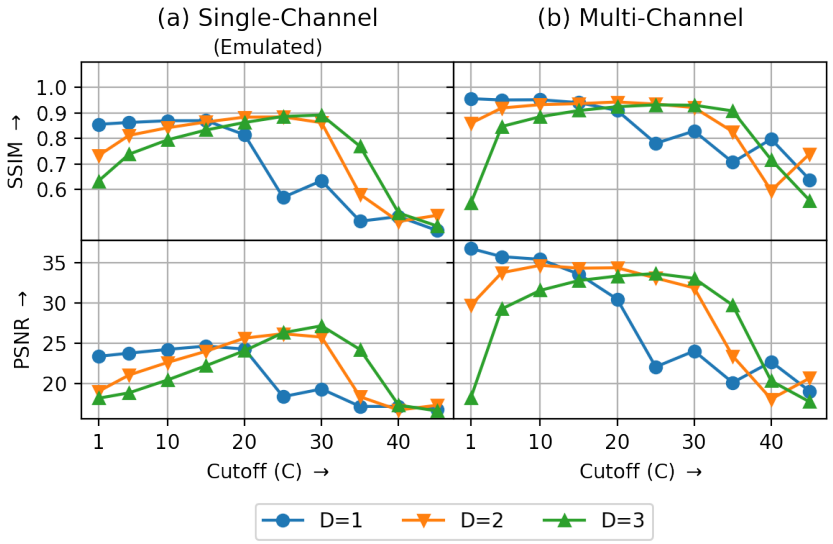


Figure A-4 – Grid search performed on the density parameters (C, D) in 2D and at UF=2.25 to obtain optimal densities in the single channel **(a)** and multi-channel **(b)** coil settings.

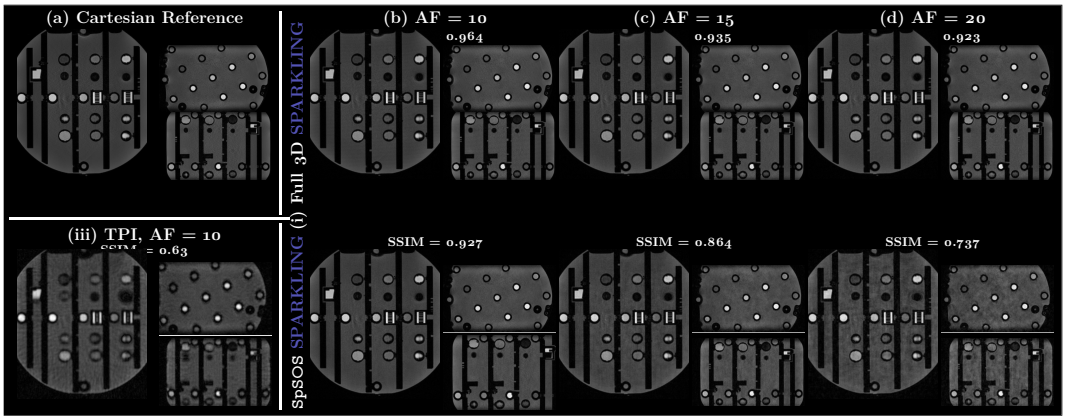


Figure A-5 – Comparison of retrospective results for **(i)** fully optimized 3D SPARKLING (right, top row) and **(ii)** SpSOS (right, bottom row) for varying acceleration factors (from left to right, AF=10 (b), 15 (c) and 20 (d)) on the NIST phantom. Additionally, we present the results of a retrospective study based on **(iii)** TPI at AF=10 for comparison purposes with the state of the art (left, bottom row). The Cartesian reference image (AF=4, GRAPPA reconstructed) is shown in (a, left top row) [Gri+02]. SSIM scores are reported for each setup. Global 3D SPARKLING gives improved results compared to the SpSOS approach which starts to get worse at AF=15 with some blurring and at AF=20 the image gets noisier. On the other hand, TPI images are extremely blurry even at AF=10.

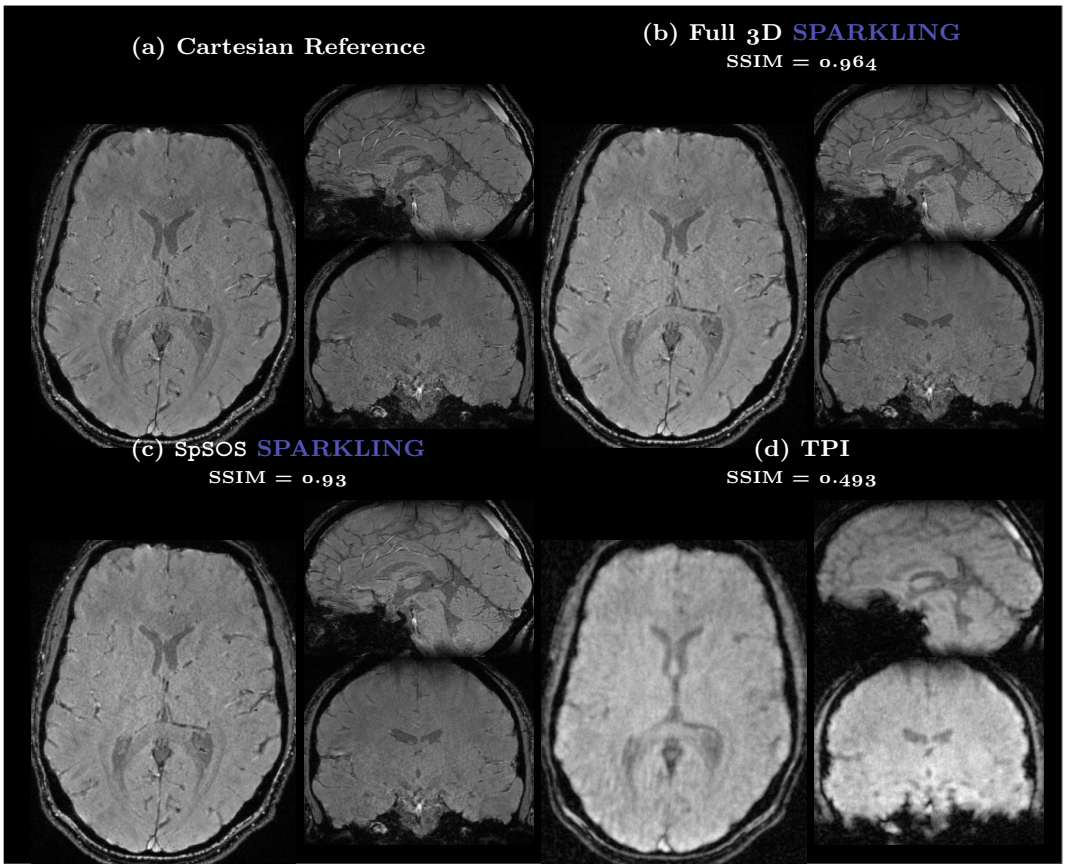


Figure A-6 – Comparing the performance of full 3D SPARKLING (b) and SpSOS SPARKLING (c) with twisted projection imaging (TPI) (d) using a retrospective study at AF=10 from the Cartesian GRAPPA-4 reference (a).

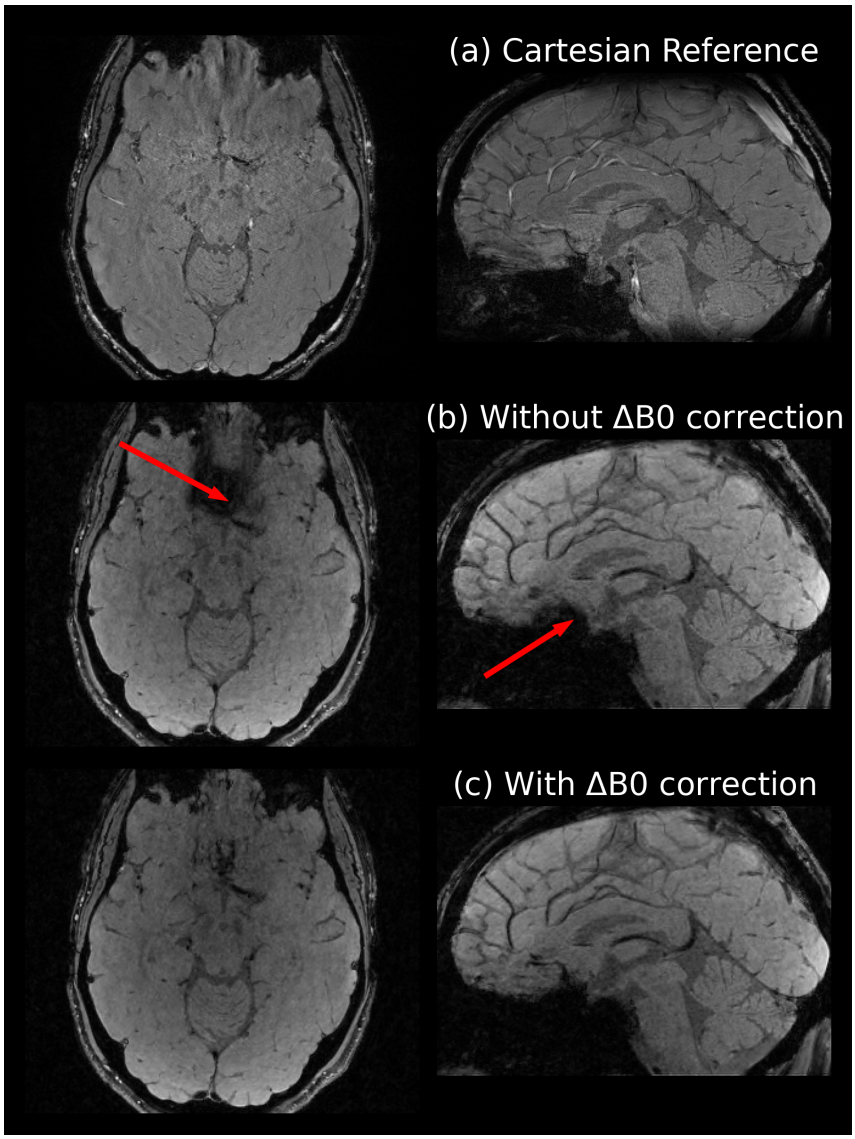


Figure A-7 – Prospective reconstruction results (axial and sagittal view only) (b) without $\Delta B0(\mathbf{r})$ correction and (c) with $\Delta B0(\mathbf{r})$ correction for full 3D SPARKLING trajectory at AF=10. Cartesian reference views (a) are also shown for comparison purpose. Red arrows in (b) refer to the regions of strong $\Delta B0$ artifacts. We see that most of the MR signal in these areas is recovered in (c) using the approach proposed in [Dav+21].

Appendix B

MORE and GoLF

B.1 Generalized TSD characterization for GoLF

In this section, we will complete mathematical details for the generic case (any $v > 0$) of TSD characterization for GoLF-Spreading Projection Algorithm for Rapid K-space sampLING (SPARKLING) . From Sec. 5.3.3, the TSD at the center of k-space will be different from the edges where we transition from Cartesian to Non-Cartesian (NC) trajectories and hence discontinuous. Thus, we need to re-paramterize the TSD into the following non-continuous form:

$$\Pi_{r_{\mathcal{S}},D}(x) = \begin{cases} \kappa_1 & |x| < r_{\mathcal{S}} \\ \kappa_2 \left(\frac{r_{\mathcal{S}}}{|x|}\right)^D & |x| > r_{\mathcal{S}} \end{cases} \tag{B.1}$$

Now, if $N = N_c \times N_s$ is the total number of samples, the number of samples in center of 3D k-space $N_{|x|<r_{\mathcal{S}}}$ is given by following volume integral in a spherical coordinate system:

$$\begin{aligned} N_{|x|<r_{\mathcal{S}}} &= N \int_0^{r_{\mathcal{S}}} \int_0^{2\pi} \int_0^{\pi} \kappa_1 x^2 \sin\theta d\theta d\phi dx \\ &= N\kappa_1 \frac{4}{3}\pi r_{\mathcal{S}}^3 \end{aligned} \tag{B.2}$$

Additionally, the number of Nyquist points in the center of k-space is given by:

$$N_{\text{Nyq}} = \frac{\frac{4}{3}\pi r_{\mathcal{S}}^3}{\Delta x^3} \tag{B.3}$$

If the trajectory velocity is given by v, we have $\frac{N_{\text{Nyq}}}{v}$ k-space samples in the center of k-space. From Eq.(B.3) and (B.2), we obtain:

$$\begin{aligned} N_{|x|<r_{\mathcal{S}}} &= \frac{N_{\text{Nyq}}}{v} \\ N\kappa_1 \frac{4}{3}\pi r_{\mathcal{S}}^3 &= \frac{\frac{4}{3}\pi r_{\mathcal{S}}^3}{v\Delta x^3} \\ \kappa_1 &= \frac{1}{vN\Delta x^3} \end{aligned} \tag{B.4}$$

Now as we switch from Cartesian sampling to NC sampling, we move to Poisson disk sampling for the NC region. From Eq. (5.15) and as the density needed for Cartesian sampling is $v \times \kappa_1$, we get:

$$\kappa_2 = v\kappa_1 \frac{N^{nc}}{N^c} = v\kappa_1 \frac{6}{\pi} \tag{B.5}$$

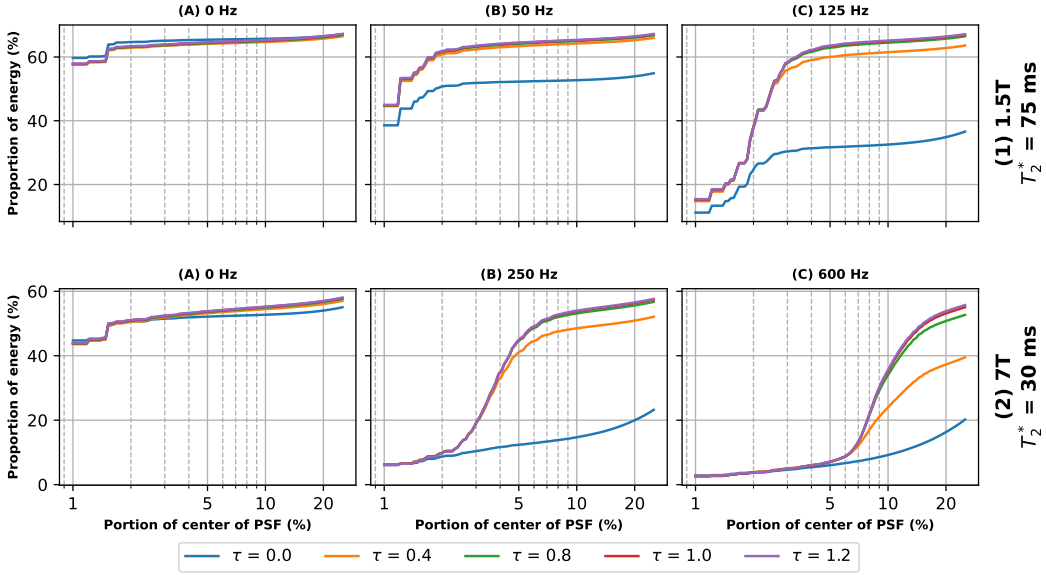


Figure B-1 – **Proportion of total energy around the center of Point Spread Function (PSF) for varying levels of ΔB_0 for 1.5T and 7T.**
The accumulated proportion of energy for Minimized Off Resonance Effect (MORE)-SPARKLING trajectories with varying τ from 0 to 1.2. The energy of the 3D PSF within a radial shell with radii at different distances from center shown on x-axis on a log scale. The PSF is obtained with a constant $T_2^*$of **(1)** 75 ms for 1.5T and **(2)** 30 ms for 7T and with off resonance frequencies of **(A)** 0 Hz, **(B)** 100 Hz and **(C)** 250 Hz.

However, in practice we cannot achieve Poisson disk sampling due to curve constraints on trajectory on speed and acceleration of the trajectory (see Sec. 7.2.7). To prevent any k-space holes, we sample the annular region where we shift from Cartesian to NC sampling at 1.5 times the Nyquist criteria, thereby choosing $\kappa_2 = v\kappa_1 \frac{N^{nc}}{N^c} = v\kappa_1 \frac{9}{\pi}$.

Finally, as $\Pi_{r_{\mathcal{S}},D}(x)$ is a distribution, we need to ensure that it is normalized, (i.e. sum

to 1):

$$\int_0^1 \int_0^{2\pi} \int_0^{\pi} \Pi_{r_S,D}(x) x^2 \sin\theta d\theta d\phi dx = 1$$
$$4\pi \int_0^1 \Pi_{r_S,D}(x) x^2 dx = 1$$
$$4\pi \left(\int_0^{r_S} \kappa_1 x^2 dx + \int_{r_S}^1 \kappa_2 \left(\frac{r_S}{|x|} \right)^D x^2 dx \right) = 1$$
$$4\pi \left(\frac{\kappa_1 r_S^3}{3} + \kappa_2 \frac{(r_S)^D - r_S^3}{3 - D} \right) = 1 \tag{B.6}$$

From Eq. (B.4), (B.5) and (B.6), we can solve iteratively for D.

B.2 Calculating PSF at 1.5 and 7T

We describe how the PSF is computed for any trajectory $\mathbf{K}$ where α and ω are respectively the T_2^* decay and off-resonance angular frequency that are used over the whole FOV. For this, we need to simulate the distortions observed during sampling of k-space when we are measuring a Dirac impulse function centered in the image domain. This Dirac delta function corresponds to a constant magnitude in k-space and following Eq. (5.5) we simulate the k-space data $\mathbf{Y} = (\mathbf{y}_i)_{i=1}^{N_c}$ acquired in presence of T_2^* decay and off-resonance as follows, $\mathbf{y}_i(t) = e^{-(\alpha+\imath\omega)t}$.

We obtain the PSF by carrying out reconstruction of this simulated k-space data as described in Sec. 5.3.6. However, we do not enforce any sparsity in wavelet domain to ensure that these sparsity based regularization does not influence the final reconstructed PSF.

With this, we present the proportion of energies in PSF at center for T_2^*-w imaging at 1.5T in Fig. B-1(1) and 7T in Fig. B-1(2). We again choose average of the T_2^* values between white and gray matter from [Pet+07], and the levels of off-resonance frequencies ω is chosen based on typically observed values at respective field strengths.

We see similar trends as observed in 3T, with a very strong blurring of the PSF peak at higher ω. Further, we still see that $\tau = 1.0$ is optimal in both the scenarios as any further increase in temporal weighting leads to minor incremental improvements to the PSF.

B.3 Maximum k-space velocity v under hardware constraints

The k-space velocity in center of k-space along readout direction is parameterized by a dimensionless parameter v, and is limited by the hardware constraints of the scanner. Particularly, following the notation set up in the core manuscript, under a maximum gradient strength $G_{\max}$ and for an image size $\tilde{N}$, we get the maximum k-space step in Ω with time Δt (i.e. $\alpha\Delta t$ in Eq. (7.1)) as:

$$\alpha\Delta t = \frac{\gamma G_{\max}}{K^x_{\max}} \Delta t = \frac{\gamma G_{\max}}{\frac{\tilde{N}}{2\mathrm{FOV}^x}} \Delta t \tag{B.7}$$

where $K^x_{\max}$ is the maximum k-space step in the readout direction x. Now as v represents the number of Nyquist voxels of size $\Delta x = \frac{2}{\tilde{N}}$ taken by trajectory in center of k-space in Δt, we get maximum velocity $v_{\max}$ as:

$$v_{\max} = \frac{\alpha\Delta t}{\Delta x} = \gamma G_{\max} \mathrm{FOV}^x \Delta t\,. \tag{B.8}$$

Using the values from Sec. 5.3.5 (i.e. $\gamma = 42.58$Mhz/T, $G_{\max} = 40$mT/m, $\mathrm{FOV}^{\mathrm{x}} = 0.23$m, $\Delta t = 10$µs), we get $v_{\max} = 3.92$.

Appendix C

Software and Open Source Contributions

One major aspect of my thesis included writing codes, which game me a lot of joy. Most of these codes are maintained on public or sometimes private repositories on GitHub with continuous integration and testing, to ensure reproducibility. Further, in the spirit of open science, I have contributed to many and also maintained some open source projects. In this section I will briefly review some of the contributions I have made through the packages I developed and maintained and also some open source contributions.

Most of these projects are based in Python, and particularly for the machine learning part, I have used the TensorFlow framework, purely for carrying forward some legacy codes from previous Ph.D. students. Some works are also in CUDA and C++ with bindings to Python to help gain speed.

C.1 Packages I wrote or maintained

C.2 SPARKLING

One of the major contributions of my thesis include formalizing the earlier works by Dr. Carole Lazarus on the SPARKLING method, through a generalized python package which works for both 2D and 3D. The largest contribution in this package is the binding of Fast Multipole Methods (FMM) and parallel implementation of the algorithm on multiple cores or GPUs. The computation times for the SPARKLING trajectories was drastically reduced from 1day to 10 minutes for 2D trajectories and nearly 3 weeks to 6 hours for 3D trajectories. This allowed us to actively iterate, build and test new trajectories for analyzing both prospective and retrospective reconstruction performance. This package allows users to set a bunch of parameters and can be directly run on a cluster or a single machine. Further, it implements the projection step of the SPARKLING algorithm in a generic way in the form of simple utility function and can be directly applied to project any other trajectory being optimized, which was used by us when we learned trajectories.

My contributions of MORE-SPARKLING and GoLF-SPARKLING are also added into the same package with addition of simple parameters that control the MORE and GoLF feature, to prevent code redundancy. Due to the presence of patents for SPARKLING , and also MORE-SPARKLING and GoLF-SPARKLING , this package is not open source and is only available to the collaborators of the project. However, we have setup processes in place to allow for research teams to gain access to the codes if a non-disclosure agreement is signed, which prevents further sharing of codes and limits its for purely research purposes.

C.3 Joint optimization of Trajectory and Reconstruction

The latter part of my thesis involved jointly learning hardware-compliant k-space sampling trajectories and reconstruction networks. For this, I have written a python package, which

includes the work on learning the sampling density and also learning the reconstruction network. This package implements a generic TensorFlow framework to learn trajectories and reconstruction networks, with cohesive modules for specifying the acquisition model, the reconstruction model and train them for a given loss function and optimizer. Further, the trajectory specifications follow the same as SPARKLING , which allows easy integration with SPARKLING package for the projection step.

For now this code is not yet open sourced, due to its dependency with the SPARKLING package, which is not open sourced. However, I plan to make this open source soon by decoupling this dependency in free time, or maintaining the dependency, but have a separate module just for projection step.

C.4 PySAP-MRI

PySAP-MRI is a plugin for the Python Sparse Data Analysis Package discussed later. This plugin was primarily focussed for MRI reconstructions and while it existed during the start of my thesis, my starting contributions involved refactoring the entire code base. After such a refactoring, I started to maintain these codes, which was actively used by teams at NeuroSpin to reconstruct prospectively acquired k-space data. Through multiple contributions to this package and also the Nonuniform Fast Fourier Transform (NUFFT) operators, the reconstruction time was reduced from 8 hours to 10 minutes for reconstructing non-Cartesian k-space data at 0.6mm isotropic resolution, from k-space data acquired on 42 channels.

C.5 Contributions

C.5.1 NUFFT operator

One major contribution during my thesis involved in the ensuring that efficient implementation of NUFFT operators were available in Python. At the start of my thesis, the only known stable and used python packages included pyNFFT [Vai+23] and PyNUFFT [Lin18] for compute on CPU and GPU respectively. However, the gpuNUFFT[Kno+14] was the most efficient implementation of NUFFT on GPU both in terms of speed and memory requirements, but was not available in Python but rather only in MATLAB. In the course of my thesis, I wrote python bindings to the gpuNUFFT library, which was then merged into mainline. Further, through implementation of concurrency on GPU, the computation time was reduced by nearly half.

As we moved to using TensorFlow models, the tensorflow-nufft [Mon22] was used. However, I contributed extensively to this project to fix the gradient computations and also prevent *NaN*s in the compute pipeline which came about as original implementation was not stable.

Estimating density compensators

As most of the work in my thesis was done in non-Cartesian Magnetic Resonance Imaging (MRI), the use of density compensators was essential to ensure faster convergence of the reconstruction and improved image quality in lesser iteration steps. Howeever, most of the libraries for NUFFT did not have methods to compute the density compensators. I proceeded to implement the density estimation algorithm proposed in [PM99] for `gpuNUFFT` [Kno+14] and also in Tensor centric implementations in `tfkbnufft` [RC23] and `torchKbNUFFT` [Muc+20].

C.5.2 Python Sparse Data Analysis Package

PySAP [Far+20a] is a software package that is the outcome of the COSMIC interdisciplinary research project (2016-2020) between the CS-MRI team at NeuroSpin and the CosmoStat laboratory, the two CEA entities where I pursued my PhD thesis. At its core, PySAP is a sparse reconstruction package that is intended to be used in multiple science contexts: astrophysics, medical imaging, non-destructive evaluation using tomographic and ultrasound imaging.

C.5.3 ModOpt

ModOpt, a module that contains the optimization algorithms used in PySAP and SPARKLING .

C.5.4 Reconstruction networks

As I actively used reconstruction networks for my thesis, I contributed to the astmri-reproducible-benchmark, which was the main contribution by Dr. Zaccharie Ramzi. Also the fastMRI [Zbo+18] data pipelines in TensorFlow was maintained in tf-fast-mri-data to which I contributed to have pipelines for learning trajectories with the network.

* * *
* *
*

Acronyms

T_{Obs} Observation time 18, 24, 32, 33, 87, 92, 95, 110, 149

ADC Analog to Digital Converter 32, 33, 42, 85, 89, 92, 111

ADMM Alternating Direction Method of Multipliers 29

AF Acceleration Factor ix, 76, 77, 78, 79, 80, 83, 90, 91, 94, 95, 96, 97, 98, 110, 114, 116, 120, 126, 151, 152

BJORK B-spline parameterized Joint Optimization of Reconstruction and K-space trajectories 60, 109, 114, 115, 116, 118, 119, 124, 151, 152, 153

CAIPIRINHA Controlled Aliasing in Parallel Imaging Results in Higher Acceleration 25

CNN Convolutional Neural Network 29, 49, 50, 51, 53

CS Compressed Sensing 1, 3, 9, 27, 28, 29, 31, 36, 41, 42, 44, 45, 52, 54, 57, 64, 65, 76, 81, 84, 87, 92, 108, 124, 132

CT Computed Tomography 53

DC Data Consistency 109, 111, 114

DCp Density Compensation 38, 41

DFT Discrete Fourier Transform 53

DL Deep Learning viii, 3, 9, 29, 44, 45, 46, 47, 48, 49, 51, 52, 53, 54, 56, 150

EPI Echo Planar Imaging 23, 24, 125

FFT Fast Fourier Transform 53, 55, 59

FID Free Induction Decay 16, 17, 18, 21, 149

FISTA Faster ISTA 29, 30, 31, 133

FLORET Fermat Looped ORthogonal Encoded Trajectories 36, 37, 150

FMM Fast Multipole Methods 69, 123, 141

fMRI Functional MRI 24, 125

FOV field-of-view 15, 18, 20, 21, 23, 32, 85, 92, 149

FWHM Full Width at Half Maximum 42, 43, 75, 123, 150, 155

GoLF Gridding of Low Frequencies viii, ix, 2, 4, 6, 7, 9, 83, 84, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 123, 124, 125, 126, 137, 141, 151, 152

GPU Graphical Processing Unit 46, 47, 52, 59

GRAPPA Generalized Autocalibrating Partially Parallel Acquisitions 53

IFFT Inverse Fast Fourier Transform 19, 20, 22, 37, 125

ISTA Iterative Soft Thresholding Algorithm 29, 31

KKT Karush-Kuhn-Tucker 113

LOUPE Learning-based Optimization of the Under-sampling PattErn 59, 124, 150

LReLU Leaky ReLU 48

MLP Multi-Layer Perceptron 49, 53

MORE Minimized Off Resonance Effect viii, 2, 4, 6, 7, 9, 83, 84, 86, 87, 91, 92, 93, 94, 95, 96, 98, 123, 124, 125, 126, 138, 141, 152

MR Magnetic Resonance 3, 9, 15, 19, 20, 21, 22, 23, 24, 25, 29, 31, 35, 36, 40, 41, 44, 45, 52, 53, 54, 55, 57, 149, 150

MRI Magnetic Resonance Imaging 1, 3, 4, 8, 9, 15, 16, 18, 20, 23, 24, 27, 28, 29, 31, 36, 38, 41, 44, 46, 52, 53, 54, 56, 57, 83, 95, 96, 108, 109, 123, 124, 125, 142

MRSI Magnetic Resonance Spectroscopy Imaging 36

MSE Mean Squared Error 27, 43, 44

MTF Modulation Transfer Function 40

NC Non-Cartesian 83, 84, 91, 137, 138

NMR Nuclear Magnetic Resonance 16, 32

NMSE Normalized Mean Squared Error 43

NUFFT Nonuniform Fast Fourier Transform 3, 8, 37, 38, 40, 41, 43, 55, 58, 59, 74, 108, 120, 125, 126, 133, 142

OASIS Open Access Series of Imaging Studies 56

PD Primal-Dual 55

PDHG Primal-Dual Hybrid Gradient 29

PET Positron Emission Tomography 53

PI Parallel Imaging 21

PILOT Physics-informed learned optimal trajectories 59, 60, 108, 109, 114, 115, 116, 117, 119, 124, 151, 152, 153

PNL Peak-to-Noise Level 42, 43, 75, 123, 150, 155

PNS Peripheral Nerve Stimulation 35, 36, 126

POGM' Proximal Optimal Gradient Method 29, 31

PReLU Parametric ReLU 48

PROJeCTOR PROjection for Jointly lEarning non-Cartesian Trajectories while Optimizing Reconstructor 3, 8, 108, 109, 115, 116, 117, 118, 119, 120, 121, 122, 124, 125, 126, 127, 152, 153

PSF Point Spread Function 28, 40, 41, 42, 43, 64, 74, 75, 76, 80, 83, 84, 92, 93, 123, 131, 132, 138, 139, 150, 151, 153, 154, 155

PSL Peak-to-Sidelobe Level 42, 43, 75, 123, 150, 155

PSNR Peak Signal-to-Noise Ratio 27, 44, 115, 118, 119, 120, 122, 124, 153

ReLU Rectified Linear Unit 48

RF Radio Frequency 15, 16, 17, 18, 19, 20, 21, 23, 32, 42, 149

RMSE Root Mean Squared Error 43, 44

SGD Stochastic Gradient Descent 47

SNR Signal-to-Noise Ratio 1, 5, 15, 20, 21, 32, 36, 44, 119

SPARKLING Spreading Projection Algorithm for Rapid K-space sampLING viii, ix, 1, 2, 3, 4, 5, 6, 7, 8, 9, 37, 59, 60, 63, 64, 65, 66, 71, 73, 74, 75, 76, 77, 78, 79, 80, 81, 83, 84, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 107, 108, 109, 114, 115, 116, 117, 118, 119, 121, 122, 123, 124, 125, 126, 131, 132, 133, 135, 136, 137, 138, 141, 142, 143, 151, 152, 153, 154

SSIM Structural Similarity Index Measure 27, 44, 115, 118, 119, 120, 122, 124, 153

SVD Singular Value Decomposition 21

SWI Susceptibility Weighted Imaging 23, 24, 123

TE Echo Time 2, 6, 35, 36, 40, 67, 72, 84, 85, 86, 87, 88, 92, 95, 96, 110, 111, 124

TPI Twisted Projection Imaging 36, 125

TPSF Transform Point Spread Function 42

TR Repetition Time 19, 23, 72, 92

TSD Target Sampling Density viii, ix, 1, 2, 3, 4, 5, 7, 8, 9, 36, 57, 59, 83, 84, 86, 87, 90, 91, 96, 108, 123, 124, 125, 131, 134, 137

TWIRL TWisting Radial Lines 36

UF Undersampling Factor 114, 118, 153

UTE Ultrashort Echo Time 36

VDS Variable Density Sampling 1, 2, 3, 5, 6, 9, 25, 27, 31, 36, 37, 44, 64, 84, 87, 108

List of Figures

List of Tables

Bibliography

[08] *e-MRI courses by IMAIOS*. https://www.imaios.com/en/e-Courses/e-MRI. Accessed: 2021-10-08. 2008 (page 16).

[18] *NHS: How it's performed - MRI scan*. https://www.nhs.uk/conditions/mri-scan/what-happens/. Accessed: 2021-10-11. 2018 (page 23).

[21] *NIST/NIBIB Medical Imaging Phantom Lending Library*. Oct. 2021 (page 92).

[Adc+13] B. Adcock, A. C. Hansen, C. Poon, B. Roman et al. "Breaking the coherence barrier: asymptotic incoherence and asymptotic sparsity in compressed sensing". In: *arXiv preprint arXiv:1302.0561* (2013) (page 108).

[Adc+17] B. Adcock et al. "Breaking the coherence barrier: A new theory for compressed sensing". In: *Forum of Math., Sigma*. Vol. 5. Cambridge University Press. 2017 (pages 1, 5, 31, 36, 64, 100).

[Ahn+86a] C. Ahn et al. "High-speed spiral-scan echo planar NMR imaging-I". In: *IEEE Trans. Med. Imag.* 5.1 (1986), pp. 2–7 (pages 1, 5, 64, 84, 100).

[Ahn+86b] C. Ahn et al. "High-speed spiral-scan echo planar NMR imaging-I". In: *IEEE Trans. Med. Imag.* 5.1 (1986), pp. 2–7 (pages 37, 108).

[AJ20] H. K. Aggarwal and M. Jacob. "J-MoDL: Joint Model-Based Deep Learning for Optimized Sampling and Reconstruction". In: *IEEE Journal of Selected Topics in Signal Processing* 14.6 (Oct. 2020), pp. 1151–1162 (page 58).

[Akç+19] M. Akçakaya, S. Moeller, S. Weingärtner and K. Uğurbil. "Scan-specific robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction: Database-free deep learning for fast imaging". In: *Magnetic resonance in medicine* 81.1 (2019), pp. 439–453 (page 53).

[AMJ18] H. K. Aggarwal, M. P. Mani and M. Jacob. "MoDL: Model-based deep learning architecture for inverse problems". In: *IEEE transactions on medical imaging* 38.2 (2018), pp. 394–405 (pages 29, 55).

[AMJ19] H. K. Aggarwal, M. P. Mani and M. Jacob. "MoDL: Model-Based Deep Learning Architecture for Inverse Problems". In: *IEEE Transactions on Medical Imaging* 38.2 (2019), pp. 394–405 (page 54).

[Amo+22a] Z. Amor, **Chaithya, G R**, B. Daval-Frérot G. Thirion, F. Mauconduit, P. Ciuciu and A. Vignaud. "3D SPARKLING for functional MRI: A pilot study for retinotopic mapping at 7T". In: *30th proceedings of ISMRM*. Glasgow, Scotland, UK, 2022 (pages 12, 125).

[Amo+22b] Z. Amor, **Chaithya, G R**, B. Daval-Frérot G. Thirion, F. Mauconduit, C. Mirkes, P. Ciuciu and A. Vignaud. "Prospects of non-Cartesian 3D-SPARKLING encoding for functional MRI: A preliminary case study for retinotopic mapping". In: *ISMRM*. 2823. London, UK, May 2022 (pages 12, 125).

[Amo+22c] Z. Amor, **Chaithya, G R**, C. Le Ster, G. Daval-Frérot, N. Boulant, F. Mauconduit, C. Mirkes, P. Ciuciu and A. Vignaud. "B_0 field distortions monitoring and correction for 3D non-Cartesian fMRI acquisitions using a field camera: Application to 3D-SPARKLING at 7T". In: *ISMRM*. 2822. London, UK, May 2022 (pages 12, 125).

[AO17] J. Adler and O. Oktem. “Solving ill-posed inverse problems using iterative deep neural networks”. In: *Inverse Problems* 33.12 (2017), pp. 1–24 (page 55).

[AÖ18] J. Adler and O. Öktem. “Learned primal-dual reconstruction”. In: *IEEE transactions on medical imaging* 37.6 (2018), pp. 1322–1332 (pages 29, 55).

[AP00] E. Ahunbay and J. G. Pipe. “Rapid method for deblurring spiral MR images”. In: *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 44.3 (2000), pp. 491–494 (page 40).

[AZ17] B. Amos and J. Zico Kolter. “OptNet: Differentiable Optimization as a Layer in Neural Networks”. In: *ICML*. 2017 (page 51).

[Bah+20a] C. D. Bahadir et al. “Deep-learning-based optimization of the under-sampling pattern in MRI”. In: *IEEE Trans. Comput. Imaging* 6 (2020), pp. 1139–1152 (pages 2, 7, 8, 59, 100, 102, 104, 108, 114).

[Bah+20b] C. D. Bahadir et al. “Deep-learning-based optimization of the under-sampling pattern in MRI”. In: *IEEE Trans. Comput. Imaging* 6 (2020), pp. 1139–1152 (pages 65, 72, 124).

[Bal+16] L. Baldassarre et al. “Learning-based compressive subsampling”. In: *IEEE J. Sel. Topics Signal Process.* 10.4 (2016), pp. 809–822 (page 65).

[Bap+22a] R. Baptista, A. Vignaud, **Chaithya, G R**, G. Daval-Frérot, F. Mauconduit, M. Naudin, M. Lapert, R. Guillevin, P. Ciuciu, C. Lerman-Rabrait and F. Boumezbeur. “Evaluation of 3D SPARKLING readout for Sodium UTE MRI at ultra-high magnetic field”. In: *ISMRM*. London, UK, May 2022 (pages 12, 125).

[Bap+22b] R. P. Baptista et al. “Evaluation of 3D SPARKLING readout for Sodium UTE MRI at ultra-high magnetic field”. In: *30th Proceedings of the ISMRM society*. 1613. London, May 2022 (page 96).

[Bar+88] J. Barzilai et al. “Two-Point Step Size Gradient Methods”. In: *IMA J. Numer. Anal.* 8.1 (1988), pp. 141–148 (page 70).

[BE10] Z. Ben-Haim and Y. C. Eldar. “The Cramér-Rao Bound for Estimating a Sparse Parameter Vector”. In: *IEEE Transactions on Signal Processing* 58.6 (2010), pp. 3384–3389 (pages 1, 6, 65).

[Bil+15] B. Bilgic, B. A. Gagoski, S. F. Cauley, A. P. Fan, J. R. Polimeni, P. E. Grant, L. L. Wald and K. Setsompop. “Wave-CAIPI for highly accelerated 3D imaging”. In: *Magn. Reson. Med.* 73.6 (2015), pp. 2152–2162 (page 25).

[BKH16] J. L. Ba, J. R. Kiros and G. E. Hinton. *Layer normalization*. Tech. rep. 2016 (page 50).

[BKZ04a] M. A. Bernstein, K. F. King and X. J. Zhou. “Signal acquisition and k-space sampling”. In: *Handbook of MRI Pulse Sequences*. Elsevier, 2004, pp. 367–442 (page 16).

[BKZ04b] M. A. Bernstein, K. F. King and X. J. Zhou. *Handbook of MRI pulse sequences*. Elsevier, 2004 (page 91).

[BMK18] A. H. Barnett, J. F. Magland and L. a. Klinteberg. *A parallel non-uniform fast Fourier transform library based on an "exponential of semicircle" kernel*. 2018 (page 38).

[BN17] C. A. Baron and D. G. Nishimura. “B0 mapping using rewinding trajectories (BMART)”. In: *Magnetic Resonance in Medicine* 78.2 (2017), pp. 664–669 (page 40).

[Boa+97] F. E. Boada et al. “Fast three dimensional sodium imaging”. In: *Magn. Reson. Med.* 37.5 (1997), pp. 706–715 (pages 1, 5, 36, 64, 65, 76, 96).

[Boy+11] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein et al. "Distributed optimization and statistical learning via the alternating direction method of multipliers". In: *Foundations and Trends® in Machine learning* 3.1 (2011), pp. 1–122 (page 29).

[Boy+16] C. Boyer et al. "On the generation of sampling schemes for MRI". In: *SIAM J. Imag. Sci.* 9.4 (2016), pp. 2039–2072 (pages 1, 5, 37, 64, 66, 86).

[Boy+19] C. Boyer et al. "Compressed sensing with structured sparsity and structured acquisition". In: *Appl. Comput. Harmon. Anal.* 46.2 (2019), pp. 312–350 (pages 1, 5, 31, 36, 64, 100, 108).

[Bre+05] F. A. Breuer et al. "Controlled aliasing in parallel imaging results in higher acceleration (CAIPIRINHA) for multi-slice imaging". In: *Magnetic Resonance in Medicine* 53.3 (2005), pp. 684–691 (pages 25, 97, 125).

[Bro+14] R. W. Brown, Y.-C. N. Cheng, E. M. Haacke, M. R. Thompson and R. Venkatesan. *Magnetic resonance imaging: physical principles and sequence design*. John Wiley & Sons, 2014 (page 16).

[BS09] I. Bayram and I. W. Selesnick. "A subband adaptive iterative shrinkage/thresholding algorithm". In: *IEEE Transactions on Signal Processing* 58.3 (2009), pp. 1131–1143 (page 29).

[BT09] A. Beck and M. Teboulle. "A fast iterative shrinkage-thresholding algorithm for linear inverse problems". In: *SIAM journal on imaging sciences* 2.1 (2009), pp. 183–202 (pages 29, 30, 55).

[Bue+07] M. Buehrer, K. P. Pruessmann, P. Boesiger and S. Kozerke. "Array compression for MRI with large coil arrays". In: *Magnetic Resonance in Medicine* 57.6 (2007), pp. 1131–1139 (pages 21, 92).

[CB18] L. Chizat and F. Bach. "On the Global Convergence of Gradient Descent for Over-parameterized Models using Optimal Transport". In: *Advances in Neural Information Processing Systems (NIPS)*. Montréal, Canada, Dec. 2018 (page 65).

[CC] **Chaithya, G R** and P. Ciuciu. "Jointly learning Non-Cartesian k-space trajectories and reconstruction networks for 2D and 3D MR imaging through projection". In: *special issue on AI in MRI: Frontiers and Applications, Bioengineering* (pages 2, 3, 8, 11, 107).

[CC22] **Chaithya, G R** and P. Ciuciu. "Benchmarking learned non-Cartesian k-space trajectories and reconstruction networks". In: *ISMRM*. 3308. London, UK, May 2022 (pages 2, 8, 12, 107).

[Cha+] **Chaithya, G R**, G. Daval-Frérot, A. Massire, A. Vignaud and P. Ciuciu. "Improving SPARKLING trajectories through Minimized Off-Resonance Effects and Gridding of Low Frequencies". In: *under review MRM* (pages 2, 6, 11, 83).

[Cha+13] N. Chauffert et al. "Variable density compressed sensing in MRI. Theoretical vs heuristic sampling strategies". In: *2013 IEEE 10th International Symposium on Biomedical Imaging*. IEEE. 2013, pp. 298–301 (pages 100, 108).

[Cha+14] N. Chauffert et al. "Variable density sampling with continuous trajectories. Application to MRI". In: *SIAM J. Imag. Sci.* 7.4 (Nov. 2014), pp. 1962–1992 (pages 1, 5, 31, 36, 37, 64, 100, 108).

[Cha+16] N. Chauffert et al. "A projection algorithm for gradient waveforms design in MRI". In: *IEEE Trans. Med. Imag.* 35.9 (Sept. 2016), pp. 2026–2039 (pages 3, 8, 32, 66, 67, 70, 71, 85, 86, 109–111, 113, 114, 119, 124).

[Cha+17] N. Chauffert et al. "A projection method on measures sets". In: *Constr. Approx.* 45.1 (2017), pp. 83–111 (pages 1, 5, 37, 59, 64, 66, 69, 70, 86, 101).

[Cha+20] B. Charlier et al. "Kernel operations on the GPU, with autodiff, without memory overflows". In: *arXiv:2004.11127* (2020) (page 68).

[Cha+22a] **Chaithya, G R**, G. Daval-Frérot, A. Massire, B. Mailhe, M. Nadar, A. Vignaud and P. Ciuciu. "MORE-SPARKLING: Non-Cartesian trajectories with Minimized Off-Resonance Effects". In: *ISMRM*. 1435. London, UK, May 2022 (pages 2, 6, 12, 83).

[Cha+22b] **Chaithya, G R**, G. Daval-Frérot, A. Vignaud and P. Ciuciu. "Method and apparatus for performing accelerated Magnetic Resonance Imaging through gridded sampling at low frequencies". Patent application in process. Dec. 2022 (pages 2, 6, 11).

[Cha+22c] **Chaithya, G R**, G. Daval-Frérot, A. Vignaud and P. Ciuciu. "Method and apparatus for performing accelerated Magnetic Resonance Imaging with reduced off-resonance effect". Patent 1000506640/EP22305592.2. Patent Application: Europe N° 22305592.2. Apr. 2022 (pages 2, 6, 11).

[Cha+22d] **Chaithya, G R**, P. Weiss, A. Massire, A. Vignaud and P. Ciuciu. "Optimizing full 3D SPARKLING trajectories for high-resolution Magnetic Resonance imaging". In: *IEEE Transactions on Medical Imaging* (Mar. 2022) (pages 1, 2, 5, 6, 11, 63, 84–86, 90, 94, 100–102, 104, 108–110, 113, 114).

[Cha15] N. Chauffert. "Echantillonnage compressé le long de trajectoires physiquement plausibles en IRM". PhD thesis. Orsay, France: Université Paris-Sud, 2015 (pages 1, 5, 64).

[Che+18] H. Cherkaoui, L. El Gueddari, C. Lazarus, A. Grigis, F. Poupon, A. Vignaud, S. Farrens, J.-L. Starck and P. Ciuciu. "Analysis vs synthesis-based regularization for combined compressed sensing and parallel MRI reconstruction at 7 tesla". In: *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE. 2018, pp. 36–40 (page 29).

[Chi+16] G. Chierchia, E. Chouzenoux, P. L. Combettes and J.-C. Pesquet. *The Proximity Operator Repository. User's guide*. http://proximity-operator.net/. Accessed: 2021-10-21. 2016 (page 30).

[Çiç+16] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox and O. Ronneberger. "3D U-Net: learning dense volumetric segmentation from sparse annotation". In: *International conference on medical image computing and computer-assisted intervention*. Springer. 2016, pp. 424–432 (page 54).

[Con13] L. Condat. "A primal–dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms". In: *Journal of optimization theory and applications* 158.2 (2013), pp. 460–479 (page 29).

[CP11a] E. J. Candes and Y. Plan. "A probabilistic and RIPless theory of compressed sensing". In: *IEEE transactions on information theory* 57.11 (2011), pp. 7235–7254 (page 31).

[CP11b] A. Chambolle and T. Pock. "A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging". In: *Journal of Mathematical Imaging and Vision* 40 (2011), pp. 120–145 (page 55).

[CP11c] A. Chambolle and T. Pock. "A first-order primal-dual algorithm for convex problems with applications to imaging". In: *Journal of mathematical imaging and vision* 40.1 (2011), pp. 120–145 (page 29).

[CRC21] **Chaithya, G R**, Z. Ramzi and P. Ciuciu. "Learning the sampling density in 2D SPARKLING MRI acquisition for optimized image reconstruction". In: *29th European Signal Processing Conference (EUSIPCO)*. Dublin, Ireland, Sept. 2021, pp. 960–964 (pages 2, 7, 8, 11, 80, 99, 108, 114, 134).

[CRC22] **Chaithya, G R**, Z. Ramzi and P. Ciuciu. "Hybrid learning of Non-Cartesian k-space trajectory and MR image reconstruction networks". In: *2022 IEEE 19th international symposium on biomedical imaging (ISBI)*. Kolkata, India, Mar. 2022 (pages 2, 8, 11, 81, 84).

[Dal+04] B. M. Dale et al. “Optimal design of k-space trajectories using a multi-objective genetic algorithm”. In: *Magn. Reson. Med.* 52.4 (2004), pp. 831–841 (pages 1, 6, 37, 64).

[Dav+21] G. Daval-Frérot et al. “Off-resonance correction non-Cartesian SWI using internal field map estimation”. In: *29th Proceedings of the ISMRM society.* 3551. virtual, May 2021 (pages 76, 79, 136).

[Dav+22] G. Daval-Frérot, A. Massire, B. Mailhe, M. Nadar, A. Vignaud and P. Ciuciu. “Iterative static field map estimation for off-resonance correction in non-Cartesian susceptibility weighted imaging”. In: *Magnetic Resonance in Medicine* 88.4 (2022), pp. 1592–1607 (pages 2, 6, 84, 120, 124).

[DDD04] I. Daubechies, M. Defrise and C. De Mol. “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint”. In: *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences* 57.11 (2004), pp. 1413–1457 (pages 29, 31).

[De +08] N. De Zanche et al. “NMR probes for measuring magnetic fields and field dynamics in MR systems”. In: *Magn. Reson. Med.* 60 (2008), pp. 176–186 (pages 80, 132).

[Del+10] B. M. Delattre, R. M. Heidemann, L. A. Crowe, J.-P. Vallée and J.-N. Hyacinthe. “Spiral demystified”. In: *Magnetic Resonance Imaging* 28.6 (July 2010), pp. 862–881 (page 36).

[DHS11] J. Duchi, E. Hazan and Y. Singer. “Adaptive Subgradient Methods for Online Learning and Stochastic Optimization”. In: *Journal of Machine Learning Research* 12.61 (2011), pp. 2121–2159 (page 47).

[Don20] M. Doneva. “Mathematical Models for Magnetic Resonance Imaging Reconstruction: An Overview of the Approaches, Problems, and Future Research Areas”. In: *IEEE Signal Processing Magazine* 37.1 (2020), pp. 24–32 (page 86).

[EB01] A. D. Elster and J. H. Burdette. *Questions and answers in magnetic resonance imaging.* Mozby, 2001 (page 16).

[Edl+19] B. L. Edlow, A. Mareyam, A. Horn, J. R. Polimeni, M. D. Tisdall, J. Augustinack, J. P. Stockmann, B. R. Diamond, A. Stevens, L. S. Tirrell, R. D. Folkerth, L. L. Wald, B. Fischl and A. van der Kouwe. “7 Tesla MRI of the ex vivo human brain at 100 micron resolution”. In: *bioRxiv* (2019) (page 15).

[Ehl+19] M. Ehler et al. “Curve Based Approximation of Measures on Manifolds by Discrepancy Minimization”. In: *arXiv:1910.06124* (2019) (page 67).

[El +18a] L. El Gueddari et al. “Self-calibrating nonlinear reconstruction algorithms for variable density sampling and parallel reception MRI”. In: *10th IEEE SAM Signal Processing WS.* Sheffield, UK, July 2018, pp. 415–419 (pages 76, 103, 132, 133).

[El +18b] L. El Gueddari, C. Lazarus, H. Carrié, A. Vignaud and P. Ciuciu. “Self-calibrating nonlinear reconstruction algorithms for variable density sampling and parallel reception MRI”. In: *2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM).* IEEE. 2018, pp. 415–419 (pages 23, 29, 91, 92).

[El +19] L. El Gueddari, P. Ciuciu, E. Chouzenoux, A. Vignaud and J.-C. Pesquet. “Calibrationless oscar-based image reconstruction in compressed sensing parallel MRI”. In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019).* IEEE. 2019, pp. 1532–1536 (page 29).

[El +21a] L. El Gueddari, **Chaithya, G R**, E. Chouzenoux and P. Ciuciu. “Calibration-Less Multi-Coil Compressed Sensing Magnetic Resonance Image Reconstruction Based on OSCAR Regularization”. In: *Journal of Imaging* 7.3 (2021) (pages 11, 23).

[El +21b] L. El Gueddari, G. R. Chaithya, E. Chouzenoux and P. Ciuciu. “Calibration-less multi-coil compressed sensing Magnetic Resonance Image reconstruction based on OSCAR regularization”. In: *Journal of Imaging* 7.3 (2021), p. 58 (page 29).

[EMR07] M. Elad, P. Milanfar and R. Rubinstein. “Analysis versus synthesis in signal priors”. In: *Inverse problems* 23.3 (2007), p. 947 (page 29).

[Eo+18] T. Eo, Y. Jun, T. Kim, J. Jang, H.-J. Lee and D. Hwang. “KIKI-net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images”. In: *Magnetic resonance in medicine* 80.5 (2018), pp. 2188–2201 (pages 54, 55).

[Far+20a] S. Farrens et al. “PySAP: Python Sparse Data Analysis Package for multidisciplinary image processing”. In: *Astron. Comput.* 32 (2020), p. 100402 (pages 76, 92, 133, 142).

[Far+20b] S. Farrens, A. Grigis, L. El Gueddari, Z. Ramzi, **Chaithya, G R**, S. Starck, B. Sarthou, H. Cherkaoui, P. Ciuciu and J.-L. Starck. “PySAP: Python Sparse Data Analysis Package for multidisciplinary image processing”. In: *Astronomy and Computing* 32 (2020), p. 100402 (page 11).

[Fei+86] D. A. Feinberg, J. D. Hale, J. C. Watts, L. Kaufman and A. Mark. “Halving MR imaging time by conjugation: demonstration at 3.5 kG.” In: *Radiology* 161.2 (1986). PMID: 3763926, pp. 527–531 (page 24).

[Fer+13] P. F. Ferreira, P. D. Gatehouse, R. H. Mohiaddin and D. N. Firmin. “Cardiovascular magnetic resonance artefacts”. In: *Journal of Cardiovascular Magnetic Resonance* 15.1 (May 2013) (page 24).

[Fes+05] J. A. Fessler, S. Lee, V. T. Olafsson, H. R. Shi and D. C. Noll. “Toeplitz-based iterative image reconstruction for MRI with correction for magnetic field inhomogeneity”. In: *IEEE Transactions on Signal Processing* 53.9 (2005), pp. 3393–3402 (page 40).

[Fes10] J. A. Fessler. “Model-Based Image Reconstruction for MRI”. In: *IEEE Signal Processing Magazine* 27.4 (2010), pp. 81–89 (page 86).

[Fes20] J. A. Fessler. “Optimization methods for magnetic resonance image reconstruction: Key models and optimization algorithms”. In: *IEEE signal processing magazine* 37.1 (2020), pp. 33–40 (page 29).

[Fon+09a] W. Fong et al. “The black-box fast multipole method”. In: *J. Comput. Phys.* 228 (2009) (pages 2, 6, 69).

[Fon+09b] W. Fong et al. “The black-box fast multipole method”. In: *J. Comput. Phys.* 228.23 (2009), pp. 8712–8725 (page 69).

[FR13] S. Foucart and H. Rauhut. “Sparse solutions of underdetermined systems”. In: *A Mathematical Introduction to Compressive Sensing*. Springer, 2013, pp. 41–59 (page 28).

[FS03] J. Fessler and B. Sutton. “Nonuniform fast Fourier transforms using min-max interpolation”. In: *IEEE Transactions on Signal Processing* 51.2 (2003), pp. 560–574 (pages 38, 111).

[GGW22] A. Gossard, F. de Gournay and P. Weiss. “Bayesian Optimization of Sampling Densities in MRI”. In: 2022 (page 59).

[GHN06] P. T. Gurney, B. A. Hargreaves and D. G. Nishimura. “Design and analysis of a practical 3D cones trajectory”. In: *Magnetic Resonance in Medicine* 55.3 (2006), pp. 575–582 (page 36).

[GJ16] R. A. de Graaf and C. Juchem. “B0 Shimming Technology”. In: *Magnetic Resonance Technology: Hardware and System Component Design*. 2016, pp. 166–207 (page 40).

[GL04] L. Greengard and J.-Y. Lee. "Accelerating the Nonuniform Fast Fourier Transform". In: *SIAM Review* 46.3 (Jan. 2004), pp. 443–454 (page 38).

[Glo99] G. H. Glover. "Simple analytic spiral k-space algorithm". In: *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 42.2 (1999), pp. 412–415 (page 36).

[Gou+21] D. Gournay et al. "Spurious minimizers in non uniform Fourier sampling optimization". working paper or preprint. June 2021 (pages 60, 65, 112).

[GOW19] D. Gilton, G. Ongie and R. Willett. "Neumann networks for linear inverse problems in imaging". In: *IEEE Transactions on Computational Imaging* 6 (2019), pp. 328–343 (pages 29, 55).

[Göz+18a] B. Gözcü et al. "Learning-based compressive MRI". In: *IEEE Trans. Med. Imag.* 37.6 (2018), pp. 1394–1406 (page 65).

[Göz+18b] B. Gözcü, R. K. Mahabadi, Y.-H. Li, E. Ilıcak, T. Çukur, J. Scarlett and V. Cevher. *Learning-Based Compressive MRI*. 2018 (page 57).

[Grä+12] M. Gräf et al. "Quadrature errors, discrepancies, and their relations to halftoning on the torus and the sphere". In: *SIAM J. Sci. Comput.* 34.5 (2012), A2760–A2791 (page 67).

[Gri+02] M. A. Griswold et al. "Generalized autocalibrating partially parallel acquisitions (GRAPPA)". In: *Magn. Reson. Med.* 47.6 (2002), pp. 1202–1210 (pages 24, 72, 76, 84, 97, 125, 135).

[Gue+20] L. E. Gueddari, **Chaithya, G R**, Z. Ramzi, S. Farrens, S. Starck, A. Grigis, J.-L. Starck and P. Ciuciu. "PySAP-MRI: a Python Package for MR Image Reconstruction". In: *ISMRM workshop on Data Sampling and Image Reconstruction*. Sedona, AZ, United States, Jan. 2020 (pages 12, 76, 133).

[Ham+19] K. Hammernik, J. Schlemper, C. Qin, J. Duan, R. M. Summers and D. Rueckert. "Sigma-net: Systematic Evaluation of Iterative Deep Neural Networks for Fast Parallel MR Image Reconstruction". In: *arXiv preprint arXiv:1912.09278* (2019) (pages 29, 55).

[Han+18] Y. Han, J. Yoo, H. H. Kim, H. J. Shin, K. Sung and J. C. Ye. "Deep learning with domain adaptation for accelerated projection-reconstruction MR". In: *Magnetic resonance in medicine* 80.3 (2018), pp. 1189–1205 (page 53).

[Han19] B. Hanin. "Universal function approximation by deep neural nets with bounded width and relu activations". In: *Mathematics* 7.10 (2019), p. 992 (page 51).

[He+15] K. He, X. Zhang, S. Ren and J. Sun. "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification". In: *ICCV*. 2015 (page 48).

[Her+02] M. A. Herráez, D. R. Burton, M. J. Lalor and M. A. Gdeisat. "Fast two-dimensional phase-unwrapping algorithm based on sorting by reliability following a noncontinuous path". In: *Applied optics* 41.35 (2002), pp. 7437–7444 (page 92).

[HFS17] J. Hamilton, D. Franson and N. Seiberlich. "Recent advances in parallel imaging for MRI". In: *Progress in Nuclear Magnetic Resonance Spectroscopy* 101 (Aug. 2017), pp. 71–95 (page 25).

[HG00] F. Hebrank and M. Gebhardt. "SAFE-Model - A New Method for Predicting Peripheral Nerve Stimulations in MRI". In: *8th proceedings of ISMRM*. Denver, CO, USA, 2000 (pages 36, 126).

[HGS15] H. van Hasselt, A. Guez and D. Silver. *Deep Reinforcement Learning with Double Q-learning*. 2015 (page 58).

[Hir+08] Y. Hirokawa, H. Isoda, Y. S. Maetani, S. Arizono, K. Shimada and K. Togashi. "MRI Artifact Reduction and Quality Improvement in the Upper Abdomen with PROPELLER and Prospective Acquisition Correction (PACE) Technique". In: *American Journal of Roentgenology* 191.4 (Oct. 2008), pp. 1154–1158 (pages 36, 37).

[HK19] J. P. Haldar and D. Kim. "OEDIPUS: An Experiment Design Framework for Sparsity-Constrained MRI". In: *IEEE Trans. Med. Imag.* 38.7 (2019), pp. 1545–1558 (pages 1, 6, 65).

[HSW89] K. Hornik, M. Stinchcombe and H. White. "Multilayer feedforward networks are universal approximators". In: *Neural networks* 2.5 (1989), pp. 359–366 (page 51).

[HSY19] Y. Han, L. Sunwoo and J. C. Ye. "k-space deep learning for accelerated MRI". In: *IEEE transactions on medical imaging* (2019) (page 53).

[Hyu+18] C. M. Hyun, H. P. Kim, S. M. Lee, S. Lee and J. K. Seo. "Deep learning for undersampled MRI reconstruction". In: *Physics in Medicine & Biology* 63.13 (2018), p. 135007 (page 53).

[Ira+95] P. Irarrazabal et al. "Fast three dimensional magnetic resonance imaging". In: *Magn. Reson. Med.* 33.5 (1995), pp. 656–662 (pages 1, 2, 5, 6, 36, 64).

[IS15] S. Ioffe and C. Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift". In: *International conference on machine learning*. PMLR. 2015, pp. 448–456 (page 50).

[Jac+91] J. Jackson, C. Meyer, D. Nishimura and A. Macovski. "Selection of a convolution function for Fourier inversion using gridding (computerised tomography application)". In: *IEEE Transactions on Medical Imaging* 10.3 (1991), pp. 473–478 (page 38).

[Jac+92] J. I. Jackson et al. "Twisting radial lines with application to robust magnetic resonance imaging of irregular flow". In: *Magn. Reson. Med.* 25.1 (1992), pp. 128–139 (pages 1, 5, 64, 84, 108).

[Jia+13] M. Jiang, J. Jin, F. Liu, Y. Yu, L. Xia, Y. Wang and S. Crozier. "Sparsity-constrained SENSE reconstruction: an efficient implementation using a fast composite splitting algorithm". In: *Magnetic Resonance Imaging* 31.7 (2013), pp. 1218–1227 (page 29).

[JNM92] J. I. Jackson, D. G. Nishimura and A. Macovski. "Twisting radial lines with application to robust magnetic resonance imaging of irregular flow". In: *Magnetic Resonance in Medicine* 25.1 (May 1992), pp. 128–139 (page 36).

[Joh17] K. M. Johnson. "Hybrid radial-cones trajectory for accelerated MRI". In: *Magn. Reson. Med.* 77.3 (2017), pp. 1068–1081 (pages 1, 5, 36, 64).

[JUY19] K. H. Jin, M. Unser and K. M. Yi. *Self-Supervised Deep Active Accelerated MRI*. 2019 (pages 57, 58).

[KAS03] D.-h. Kim, E. Adalsteinsson and D. M. Spielman. "Simple analytic variable density spiral design". In: *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 50.1 (2003), pp. 214–219 (page 36).

[KB14] D. P. Kingma and J. Ba. "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980* (2014) (page 47).

[KF18] D. Kim and J. A. Fessler. "Adaptive restart of the optimized gradient method for convex optimization". In: *Journal of Optimization Theory and Applications* 178.1 (2018), pp. 240–263 (page 29).

[KKP09] J. Keiner, S. Kunis and D. Potts. "Using NFFT3 - a Software Library for Various Nonequispaced Fast Fourier Transforms". In: *ACM Trans. Math. Software* 36 (2009), Article 19, 1–30 (page 38).

[Kno+11] F. Knoll et al. "Adapted random sampling patterns for accelerated MRI". In: *MRM in Physics, Biology and Medicine* 24.1 (Feb. 2011), pp. 43–50 (pages 7, 57, 100, 102, 104).

[Kno+14] F. Knoll et al. "gpuNUFFT-An open source GPU library for 3D regridding with direct Matlab interface". In: *Proc. 22nd ISMRM.* 2014, p. 4297 (pages 38, 76, 133, 142).

[Kno+20] F. Knoll et al. "Advancing machine learning for MR image reconstruction with an open competition: Overview of the 2019 fastMRI challenge". In: *MRM* 84.6 (June 2020), pp. 3054–3070 (pages 7, 100).

[Kri09] A. Krizhevsky. *Learning Multiple Layers of Features from Tiny Images.* Tech. rep. 2009 (page 49).

[Kum+08] C. Kumar Anand et al. "Durga: A heuristically-optimized data collection strategy for volumetric magnetic resonance imaging". In: *Eng. Optim.* 40.2 (2008), pp. 117–136 (pages 1, 6, 37, 64).

[LaM+18] P. J. LaMontagne, S. Keefe, W. Lauren, C. Xiong, E. A. Grant, K. L. Moulder, J. C. Morris, T. L. Benzinger and D. S. Marcus. "OASIS-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer's disease". In: *Alzheimer's and Dementia: The Journal of the Alzheimer's Association* 14.7 (2018), P1097 (page 56).

[Lar+08] P. E. Larson et al. "Anisotropic field-of-views in radial imaging". In: *IEEE Trans. Med. Imag.* 27.1 (2008), pp. 47–57 (pages 1, 5, 36, 64, 125).

[Lau73] P. C. Lauterbur. "Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance". In: *Nature* 242.5394 (Mar. 1973), pp. 190–191 (pages 36, 37, 100).

[Law+09] C. S. Law et al. "Interleaved spiral-in/out with application to functional MRI (fMRI)". In: *Magn. Reson. Med.* 62.3 (2009), pp. 829–834 (pages 1, 5, 64, 84, 108).

[Laz+19] C. Lazarus et al. "SPARKLING: variable-density k-space filling curves for accelerated T2*-weighted MRI". In: *Magn. Reson. Med.* 81.6 (2019), pp. 3643–3661 (pages 1, 5, 37, 59, 60, 64, 66, 67, 73, 84, 86, 87, 100, 108, 111).

[Laz+20a] C. Lazarus et al. "3D variable-density SPARKLING trajectories for high-resolution T2*-weighted magnetic resonance imaging". In: *NMR Biomed.* (2020), e4349 (pages 2, 6, 64, 65, 79, 100, 108).

[Laz+20b] C. Lazarus et al. "Correcting the side effects of ADC filtering in MR image reconstruction". In: *J. Math. Imaging Vis.* (2020), pp. 1–14 (page 67).

[Laz18] C. Lazarus. "Compressed Sensing in MRI : optimization-based design of k-space filling curves for accelerated MRI". PhD thesis. 2018 (page 16).

[LDP07] M. Lustig, D. Donoho and J. M. Pauly. "Sparse MRI: The application of compressed sensing for rapid MR imaging". In: *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 58.6 (2007), pp. 1182–1195 (pages 28, 29).

[Leb+19] L. Lebrat et al. "Optimal Transport Approximation of 2-Dimensional Measures". In: *SIAM J. Imag. Sci.* 12.2 (Jan. 2019), pp. 762–787 (pages 3, 8, 71, 109, 113).

[LeC+89] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel. "Backpropagation applied to handwritten zip code recognition". In: *Neural computation* 1.4 (1989), pp. 541–551 (page 49).

[Lee+03] J. H. Lee, B. A. Hargreaves, B. S. Hu and D. G. Nishimura. "Fast 3D imaging using variable-density spiral trajectories with applications to limb perfusion". In: *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 50.6 (2003), pp. 1276–1285 (page 36).

[Lee+18] D. Lee, J. Yoo, S. Tak and J. C. Ye. "Deep residual learning for accelerated MRI using magnitude and phase networks". In: *IEEE Transactions on Biomedical Engineering* 65.9 (2018), pp. 1985–1995 (page 53).

[Lin+12] W. Lin, F. Huang, E. Simonotto, G. R. Duensing and A. Reykowski. "Off-resonance artifacts correction with convolution in k-space (ORACLE)". In: *Magnetic Resonance in Medicine* 67.6 (2012), pp. 1547–1555 (page 40).

[Lin18] J.-M. Lin. "Python Non-Uniform Fast Fourier Transform (PyNUFFT): An Accelerated Non-Cartesian MRI Package on a Heterogeneous Platform (CPU/GPU)". In: *Journal of Imaging* 4.3 (Mar. 2018), p. 51 (pages 38, 142).

[Liu+12] D.-d. Liu et al. "Under-sampling trajectory design for compressed sensing MRI". In: *Conf. Proc. IEEE Eng. Med. Biol. Soc.* IEEE. 2012, pp. 73–76 (page 37).

[Liu+19] L. Liu, H. Jiang, P. He, W. Chen, X. Liu, J. Gao and J. Han. "On the variance of the adaptive learning rate and beyond". In: *arXiv preprint arXiv:1908.03265* (2019) (page 47).

[Loo8] W. Lin and other. "Respiratory motion-compensated radial dynamic contrast-enhanced (DCE)-MRI of chest and abdominal lesions". In: *Magn. Reson. Med.* 60.5 (2008), pp. 1135–1146 (pages 1, 6, 64).

[LS18] J. Liang and C.-B. Schönlieb. *Improving FISTA: Faster, Smarter and Greedier*. Tech. rep. 2018 (page 31).

[Lus+05a] M. Lustig et al. "Faster imaging with randomly perturbed, under-sampled spirals and ℓ_1 reconstruction". In: *Proc. 13th ISMRM*. Miami Beach, FL, USA, 2005, p. 685 (pages 1, 5, 64, 84, 108).

[Lus+05b] M. Lustig, J. M. Santos, J.-H. Lee, D. L. Donoho and J. M. Pauly. "Application of compressed sensing for rapid MR imaging". In: *SPARS,(Rennes, France)* (2005) (page 28).

[Lus+07] M. Lustig et al. "Sparse MRI: The application of compressed sensing for rapid MR imaging". In: *Magn. Reson. Med.* 58.6 (2007), pp. 1182–1195 (pages 1, 5, 42, 64, 100, 108).

[Mar+19] H. Maron, E. Fetaya, N. Segol and Y. Lipman. "On the universality of invariant networks". In: *International conference on machine learning*. PMLR. 2019, pp. 4363–4371 (page 51).

[McG+93] G. McGibney, M. R. Smith, S. T. Nichols and A. Crawley. "Quantitative evaluation of several partial fourier reconstruction algorithms used in mri". In: *Magnetic Resonance in Medicine* 30.1 (July 1993), pp. 51–59 (page 24).

[Mey+92] C. H. Meyer et al. "Fast spiral coronary artery imaging". In: *Magn. Reson. Med.* 28.2 (1992), pp. 202–213 (pages 1, 5, 64, 84, 108).

[MHN13] A. L. Maas, A. Y. Hannun and A. Y. Ng. "Rectifier nonlinearities improve neural network acoustic models". In: *Proc. icml*. Vol. 30. 1. 2013, p. 3 (page 48).

[Mir+04] R. Mir et al. "Fast three-dimensional k-space trajectory design using missile guidance ideas". In: *Magn. Reson. Med.* 52.2 (2004), pp. 329–336 (pages 1, 6, 37, 64).

[Mon22] J. Montalt Tordera. *TensorFlow NUFFT*. Version v0.12.0. Nov. 2022 (pages 38, 59, 114, 142).

[MPM97a] L.-C. Man, J. M. Pauly and A. Macovski. "Improved automatic off-resonance correction without a field map in spiral imaging". In: *Magnetic Resonance in Medicine* 37.6 (1997), pp. 906–913 (page 40).

[MPM97b] L.-C. Man, J. M. Pauly and A. Macovski. "Multifrequency interpolation for fast off-resonance correction". In: *Magnetic Resonance in Medicine* 37.5 (1997), pp. 785–792 (page 40).

[MSS21] Q. Merigot, F. Santambrogio and C. Sarrazin. "Non-asymptotic convergence bounds for Wasserstein approximation using point clouds". working paper or preprint. June 2021 (page 65).

[Muc+20] M. J. Muckley, R. Stern, T. Murrell and F. Knoll. "TorchKbNufft: A High-Level, Hardware-Agnostic Non-Uniform Fast Fourier Transform". In: *ISMRM Workshop on Data Sampling & Image Reconstruction*. Source code available at https://github.com/mmuckley/torchkbnufft. 2020 (pages 38, 59, 142).

[Muc+21a] M. J. Muckley et al. "Results of the 2020 fastMRI Challenge for Machine Learning MR Image Reconstruction". In: *IEEE TMI* 40.9 (Sept. 2021), pp. 2306–2317 (page 114).

[Muc+21b] M. J. Muckley et al. "Results of the 2020 fastMRI Challenge for Machine Learning MR Image Reconstruction". In: *IEEE Transactions on Medical Imaging* (2021) (pages 7, 54, 100, 102).

[Nes83] Y. E. Nesterov. "A method for solving the convex programming problem with convergence rate $O(1/k^2)$". In: *Dokl. Akad. Nauk SSSR* 269 (1983), pp. 543–547 (pages 30, 31).

[NH10] V. Nair and G. E. Hinton. "Rectified linear units improve restricted boltzmann machines". In: *Icml*. 2010 (page 48).

[NN00] K. S. Nayak and D. G. Nishimura. "Automatic field map generation and off-resonance correction for projection reconstruction imaging". In: *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 43.1 (2000), pp. 151–154 (page 40).

[Nol+91] D. C. Noll, C. H. Meyer, J. M. Pauly, D. G. Nishimura and A. Macovski. "A homogeneity correction method for magnetic resonance imaging with time-varying gradients". In: *IEEE transactions on medical imaging* 10.4 (1991), pp. 629–637 (page 40).

[Nol97] D. C. Noll. "Multishot rosette trajectories for spectrally selective MR imaging". In: *IEEE Trans. Med. Imag.* 16.4 (1997), pp. 372–377 (pages 1, 5, 36, 37, 64, 84, 108).

[Nwa+18] C. Nwankpa, W. Ijomah, A. Gachagan and S. Marshall. "Activation functions: Comparison of trends in practice and research for deep learning". In: (2018) (page 48).

[Oku+17] R. Okuta et al. "CuPy : A NumPy-Compatible Library for NVIDIA GPU Calculations". In: 2017 (page 70).

[Ozt+22] B. Ozturkler, A. Sahiner, T. Ergen, A. D. Desai, C. M. Sandino, S. Vasanawala, J. M. Pauly, M. Mardani and M. Pilanci. "GLEAM: Greedy Learning for Large-Scale Accelerated MRI Reconstruction". In: *arXiv preprint arXiv:2207.08393* (2022) (page 54).

[Pao+15] M. Paolini, D. Keeser, M. Ingrisch, N. Werner, N. Kindermann, M. Reiser and J. Blautzik. "Resting-state networks in healthy adult subjects: a comparison between a 32-element and an 8-element phased array head coil at 3.0 Tesla". In: *Acta Radiologica* 56.5 (May 2015), pp. 605–613 (page 21).

[Par+14] D. L. Parker et al. "Phase reconstruction from multiple coil data using a virtual reference coil". In: *Magnetic Resonance in Medicine* 72.2 (2014), pp. 563–569 (pages 23, 72, 109, 134).

[Pat+20] F. Patzig, B. Wilm, S. Gross, D. Brunner and K. P. Pruessmann. "Off-Resonance Self-Correction for Single-Shot Imaging". In: *Proceedings of the 28th Annual Meeting of ISMRM*. Virtual, 2020, p. 3400 (page 40).

[Pau] J. M. Pauly. *AdvRecon_Pauly_read.pdf*. https://users.fmrib.ox.ac.uk/ karla/reading_group/lecture_no (Accessed on 05/04/2021) (pages 37, 41, 75).

[Pet+07] A. M. Peters, M. J. Brookes, F. G. Hoogenraad, P. A. Gowland, S. T. Francis, P. G. Morris and R. Bowtell. "T2* measurements in human brain at 1.5, 3 and 7 T". In: *Magnetic Resonance Imaging* 25.6 (2007). Proceedings of the International School on Magnetic Resonance and Brain Function, pp. 748–753 (pages 93, 139).

[Pez+20a] N. Pezzotti et al. "An Adaptive Intelligence Algorithm for Undersampled Knee MRI Reconstruction". In: *IEEE Access* 8 (2020), pp. 204825–204838 (page 112).

[Pez+20b] N. Pezzotti, S. Yousefi, M. S. Elmahdy, J. H. F. van Gemert, C. Schuelke, M. Doneva, T. Nielsen, S. Kastryulin, B. P. Lelieveldt, M. J. van Osch, E. D. Weerdt and M. Staring. "An adaptive intelligence algorithm for undersampled knee MRI reconstruction". In: *IEEE Access* 8 (2020), pp. 204825–204838 (page 51).

[Pin+20] L. Pineda, S. Basu, A. Romero, R. Calandra and M. Drozdzal. "Active MR k-space Sampling with Reinforcement Learning". In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*. Springer International Publishing, 2020, pp. 23–33 (page 58).

[Pin21] B. Pinho Meneses. "Static field shimming in the human brain for ultra-high field MRI : conceptual limits and development of a novel hardware prototype". Theses. Université Paris-Saclay, Apr. 2021 (page 39).

[Pip+11a] J. G. Pipe, N. R. Zwart, E. A. Aboussouan, R. K. Robison, A. Devaraj and K. O. Johnson. "A new design and rationale for 3D orthogonally oversampled k-space trajectories". In: *Magnetic Resonance in Medicine* 66.5 (Apr. 2011), pp. 1303–1311 (pages 36, 37).

[Pip+11b] J. G. Pipe, N. R. Zwart, E. A. Aboussouan, R. K. Robison, A. Devaraj and K. O. Johnson. "A new design and rationale for 3D orthogonally oversampled k-space trajectories". In: *Magnetic resonance in medicine* 66.5 (2011), pp. 1303–1311 (page 96).

[Pip99] J. G. Pipe. "Motion correction with PROPELLER MRI: Application to head motion and free-breathing cardiac imaging". In: *Magnetic Resonance in Medicine* 42.5 (Nov. 1999), pp. 963–969 (pages 36, 37).

[PM99] J. G. Pipe and P. Menon. "Sampling density compensation in MRI: Rationale and an iterative numerical solution". In: *Magn. Reson. Med.* 41.1 (1999), pp. 179–186 (pages 38, 76, 103, 111, 133, 142).

[Poo+22] K. Pooja, Z. Ramzi, **Chaithya, G R** and P. Ciuciu. "MC-PDNet: Deep Unrolled Neural Network for Multi-contrast MR Image Reconstruction from Undersampled k-space data". In: *2022 IEEE 19th international symposium on biomedical imaging (ISBI)*. Kolkata, India, Mar. 2022 (page 12).

[PP04] A. Pavelka and A. Procházka. "Algorithms for initialization of neural network weights". In: *In Proceedings of the 12th annual conference, MATLAB*. 2004, pp. 453–459 (page 51).

[Pru+99] K. P. Pruessmann, M. Weiger, M. B. Scheidegger and P. Boesiger. "SENSE: Sensitivity encoding for fast MRI". In: *Magnetic Resonance in Medicine* 42.5 (1999), pp. 952–962 (pages 24, 25, 97, 125).

[Pui+21] T. Puiseux, A. Sewonu, R. Moreno, S. Mendez and F. Nicoud. "Numerical simulation of time-resolved 3D phase-contrast magnetic resonance imaging". In: *PLOS ONE* 16.3 (Mar. 2021). Ed. by I. Borazjani, e0248816 (page 19).

[Puy+11] G. Puy et al. "On variable density compressive sampling". In: *IEEE Signal Process. Lett.* 18.10 (2011), pp. 595–598 (pages 1, 5, 31, 36, 37, 64, 100, 108).

[PWP21] F. Patzig, B. Wilm and K. P. Pruessmann. "Off-Resonance Self-Correction by Implicit Bo-Encoding". In: *Proceedings of the 29th Annual Meeting of ISMRM*. Virtual, 2021, p. 666 (page 40).

[Ram+20] Z. Ramzi et al. "Benchmarking MRI Reconstruction Neural Networks on Large Public Datasets". In: *Applied Sciences* 10.5 (2020), p. 1816 (pages 55, 56).

[Ram+22a] Z. Ramzi, **Chaithya, G R**, J.-L. Starck and P. Ciuciu. "Density-Compensated Unrolled Networks for 2D and 3D non-Cartesian MRI Reconstruction". In: *IEEE Transactions on Medical Imaging* (Jan. 2022) (page 11).

[Ram+22b] Z. Ramzi et al. "NC-PDNet: a Density-Compensated Unrolled Network for 2D and 3D non-Cartesian MRI Reconstruction". In: *IEEE TMI* (2022) (pages 109, 111, 112, 114, 119, 124).

[Ram+22c] Z. Ramzi, G. Chaithya, J.-L. Starck and P. Ciuciu. "NC-PDNet: a density-compensated unrolled network for 2D and 3D non-Cartesian MRI reconstruction". In: *IEEE Transactions on Medical Imaging* (2022) (page 55).

[Ram22] Z. Ramzi. "Advanced deep neural networks for MRI image reconstruction from highly undersampled data in challenging acquisition settings". Theses. Université Paris-Saclay, Feb. 2022 (page 16).

[RAP17] R. K. Robison, A. G. Anderson III and J. G. Pipe. "Three-dimensional ultrashort echo-time imaging using a floret trajectory". In: *Magnetic Resonance in Medicine* 78.3 (2017), pp. 1038–1049 (page 96).

[Ras+99] V. Rasche, R. Proksa, R. Sinkus, P. Bornert and H. Eggers. "Resampling of data between arbitrary grids using convolution interpolation". In: *IEEE Transactions on Medical Imaging* 18.5 (May 1999), pp. 385–392 (page 38).

[Rau10] H. Rauhut. "Compressive sensing and structured random matrices". In: *Theoretical foundations and numerical methods for sparse recovery* 9.1 (2010), p. 92 (page 31).

[RB10] S. Ravishankar and Y. Bresler. "MR image reconstruction from highly undersampled k-space data by dictionary learning". In: *IEEE transactions on medical imaging* 30.5 (2010), pp. 1028–1041 (page 29).

[RB11] S. Ravishankar and Y. Bresler. "Adaptive sampling design for compressed sensing MRI". In: *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE. 2011, pp. 3751–3755 (page 37).

[RB15] S. Ravishankar and Y. Bresler. "Sparsifying transform learning with efficient optimal updates and convergence guarantees". In: *IEEE Transactions on Signal Processing* 63.9 (2015), pp. 2389–2404 (page 29).

[RC23] Z. Ramzi and **Chaithya, G R**. *GitHub - zaccharieramzi/tfkbnufft: A robust, easy-to-deploy non-uniform Fast Fourier Transform in TensorFlow*. https://github.com/zaccharieramzi/tfkbnufft. [Accessed 27-Jan-2023]. 2023 (pages 38, 59, 142).

[RCS20] Z. Ramzi, P. Ciuciu and J.-L. Starck. "Benchmarking Deep Nets MRI Reconstruction Models on the FastMRI Publicly Available Dataset". In: *2020 IEEE 17th ISBI*. Iowa City (virtual), United States, Apr. 2020, pp. 1441–1445 (pages 7, 100).

[RFB15a] O. Ronneberger, P. Fischer and T. Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. 2015 (page 102).

[RFB15b] O. Ronneberger, P. Fischer and T. Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241 (pages 53, 59).

[Rob+11] S. Robinson, G. Grabner, S. Witoszynskyj and S. Trattnig. "Combining phase images from multi-channel RF coils using 3D phase offset maps derived from a dual-echo scan". In: *Magnetic resonance in medicine* 65.6 (2011), pp. 1638–1648 (page 40).

[Rob+17] S. D. Robinson, K. Bredies, D. Khabipova, B. Dymerska, J. P. Marques and F. Schweser. "An illustrated comparison of processing methods for MR phase imaging and QSM: combining array coil signals and phase unwrapping". In: *NMR in Biomedicine* 30.4 (2017), e3601 (page 23).

[Roe+90] P. B. Roemer, W. A. Edelstein, C. E. Hayes, S. P. Souza and O. M. Mueller. "The NMR phased array". In: *Magnetic resonance in medicine* 16.2 (1990), pp. 192–225 (page 22).

[Ros58] F. Rosenblatt. "The perceptron: a probabilistic model for information storage and organization in the brain." In: *Psychological review* 65.6 (1958), p. 386 (page 49).

[RSC21a] Z. Ramzi, J.-L. Starck and P. Ciuciu. "Density Compensated Unrolled Networks for Non-Cartesian MRI Reconstruction". In: *2021 IEEE 18th ISBI*. Nice, France (virtual), Apr. 2021, pp. 1443–1447 (pages 7, 101, 103).

[RSC21b] Z. Ramzi, J.-L. Starck and P. Ciuciu. "XPDNet for MRI Reconstruction: An application to the 2020 fastMRI challenge". In: *2021 ISMRM annual meeting*. Abstract 275. May 2021 (pages 55, 102, 104).

[RZL17] P. Ramachandran, B. Zoph and Q. V. Le. "Searching for activation functions". In: *arXiv preprint arXiv:1710.05941* (2017) (page 48).

[San+18] S. Santurkar, D. Tsipras, A. Ilyas and A. Madry. "How does batch normalization help optimization?" In: *Advances in neural information processing systems* 31 (2018) (page 50).

[San+19] T. Sanchez, B. Gözcü, R. B. van Heeswijk, A. Eftekhari, E. Ilıcak, T. Çukur and V. Cevher. *Scalable Learning-Based Sampling Optimization for Compressive Dynamic MRI*. 2019 (page 57).

[SBN00] D. J. Schaefer, J. D. Bourland and J. A. Nyenhuis. "Review of Patient Safety in Time-Varying Gradient Fields". In: *Journal of Magnetic Resonance Imaging* 12.1 (2000), pp. 20–29 (page 36).

[Sch+10] C. Schmaltz et al. "Electrostatic halftoning". In: *Computer Graphics Forum*. Vol. 29. 8. Wiley Online Library. 2010, pp. 2313–2327 (page 67).

[Sch+17] J. Schlemper, J. Caballero, J. V. Hajnal, A. Price and D. Rueckert. "A deep cascade of convolutional neural networks for MR image reconstruction". In: *International conference on information processing in medical imaging*. Springer. 2017, pp. 647–658 (page 54).

[Sch+19] J. Schlemper, S. Sadegh, M. Salehi, P. Kundu, C. Lazarus, H. Dyvorne, D. Rueckert and M. Sofka. "Nonuniform Variational Network: Deep Learning for Accelerated Nonuniform MR Image Reconstruction". In: *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2019, pp. 57–64 (page 55).

[See+10] M. Seeger et al. "Optimization of k-space trajectories for compressed sensing by Bayesian experimental design". In: *Magn. Reson. Med.* 63.1 (Jan. 2010), pp. 116–126 (pages 1, 6, 37, 65).

[SF09] I. W. Selesnick and M. A. Figueiredo. "Signal restoration with overcomplete wavelet transforms: Comparison of analysis and synthesis priors". In: *Wavelets XIII*. Vol. 7446. spie. 2009, pp. 107–121 (page 29).

[She+20] F. Sherry et al. "Learning the sampling pattern for MRI". In: *IEEE Trans. Med. Imag.* (2020) (pages 7, 57, 65, 72, 80, 100).

[Shi+21] Y.-h. Shih, G. Wright, J. Andén, J. Blaschke and A. H. Barnett. *cuFINUFFT: a load-balanced GPU library for general-purpose nonuniform FFTs*. 2021 (pages 38, 114).

[Shi+22] E. Shimron, J. I. Tamir, K. Wang and M. Lustig. "Implicit data crimes: Machine learning bias arising from misuse of public data". In: *Proceedings of the National Academy of Sciences* 119.13 (Mar. 2022) (page 56).

[SNF03] B. P. Sutton, D. C. Noll and J. A. Fessler. "Fast, iterative image reconstruction for MRI in the presence of field inhomogeneities". In: *IEEE transactions on medical imaging* 22.2 (2003), pp. 178–188 (page 40).

[SNF04] B. P. Sutton, D. C. Noll and J. A. Fessler. "Dynamic field map estimation using a spiral-in/spiral-out acquisition". In: *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 51.6 (2004), pp. 1194–1204 (page 40).

[Son+04] H. K. Song et al. "Dynamic MRI with projection reconstruction and KWIC processing for simultaneous high spatial and temporal resolution". In: *Magn. Reson. Med.* 52.4 (2004), pp. 815–824 (pages 1, 6, 64).

[Sou+18] R. Souza, O. Lucena, J. Garrafa, D. Gobbi, M. Saluzzi, S. Appenzeller, L. Rittner, R. Frayne and R. Lotufo. "An open, multi-vendor, multi-field-strength brain MR dataset and analysis of publicly available skull stripping methods agreement". In: *NeuroImage* 170 (2018). Segmenting the Brain, pp. 482–494 (pages 44, 56, 110, 116).

[Sou16] C. A. de Sousa. "An overview on weight initialization methods for feedforward neural networks". In: *2016 International Joint Conference on Neural Networks (IJCNN)*. IEEE. 2016, pp. 52–59 (page 51).

[Sri+14] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov. "Dropout: a simple way to prevent neural networks from overfitting". In: *The journal of machine learning research* 15.1 (2014), pp. 1929–1958 (page 51).

[Sri+20] A. Sriram, J. Zbontar, T. Murrell, A. Defazio, C. L. Zitnick, N. Yakubova, F. Knoll and P. Johnson. "End-to-end variational networks for accelerated MRI reconstruction". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2020, pp. 64–73 (pages 29, 55).

[ST95] H. Schomberg and J. Timmer. "The gridding method for image reconstruction by Fourier transformation". In: *IEEE Transactions on Medical Imaging* 14.3 (1995), pp. 596–607 (page 38).

[STM91] M. K. Stehling, R. Turner and P. Mansfield. "Echo-Planar Imaging: Magnetic Resonance Imaging in a Fraction of a Second". In: *Science* 254.5028 (Oct. 1991), pp. 43–50 (page 23).

[Stu+15] D. Stucht, K. A. Danishad, P. Schulze, F. Godenschweger, M. Zaitsev and O. Speck. "Highest Resolution In Vivo Human Brain MRI Using Prospective Motion Correction". In: *PLOS One* 10.7 (July 2015) (page 15).

[Sut+03] B. Sutton et al. "Fast, iterative image reconstruction for MRI in the presence of field inhomogeneities". In: *IEEE Trans. Med. Imag.* 22.2 (2003), pp. 178–188 (pages 2, 6, 76, 84).

[Sut+13] I. Sutskever, J. Martens, G. Dahl and G. Hinton. "On the importance of initialization and momentum in deep learning". In: *International conference on machine learning*. PMLR. 2013, pp. 1139–1147 (page 47).

[Tel16] M. Telgarsky. "Benefits of depth in neural networks". In: *Conference on learning theory*. PMLR. 2016, pp. 1517–1539 (page 51).

[Teu+11] T. Teuber et al. "Dithering by differences of convex functions". In: *SIAM J. Imag. Sci.* 4.1 (2011), pp. 79–108 (page 67).

[TF95] G. Thimm and E. Fiesler. "Neural network initialization". In: *International Workshop on Artificial Neural Networks*. Springer. 1995, pp. 535–542 (page 51).

[The+99] D. R. Thedens et al. "Fast magnetic resonance coronary angiography with a three-dimensional stack of spirals trajectory". In: *Magn. Reson. Med.* 41.6 (1999), pp. 1170–1179 (pages 2, 6, 64).

[Tom+17] T. Tommasi, N. Patricia, B. Caputo and T. Tuytelaars. "A deeper look at dataset bias". In: *Domain adaptation in computer vision applications*. Springer, 2017, pp. 37–55 (page 52).

[Uec+13] M. Uecker, P. Lai, M. J. Murphy, P. Virtue, M. Elad, J. M. Pauly, S. S. Vasanawala and M. Lustig. "ESPIRiT-an eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA". In: *Magnetic Resonance in Medicine* 71.3 (May 2013), pp. 990–1001 (page 29).

[Uec+14] M. Uecker, P. Lai, M. J. Murphy, P. Virtue, M. Elad, J. M. Pauly, S. S. Vasanawala and M. Lustig. "ESPIRiT—an eigenvalue approach to autocalibrating parallel MRI: where SENSE meets GRAPPA". In: *Magnetic resonance in medicine* 71.3 (2014), pp. 990–1001 (page 23).

[UVL16] D. Ulyanov, A. Vedaldi and V. Lempitsky. *Instance normalization: The missing ingredient for fast stylization*. Tech. rep. 2016 (page 50).

[Vai+23] G. Vaillant et al. *pyNFFT: a pythonic wrapper around the NFFT library*. https://github.com/pyNFFT/pyNFFT. [Accessed 27-Jan-2023]. 2023 (pages 38, 142).

[Vas+10] S. S. Vasanawala et al. "Improved pediatric MR imaging with compressed sensing". In: *Radiology* 256.2 (2010), pp. 607–616 (pages 1, 5, 64).

[Ved+20] S. Vedula et al. "3D FLAT: Feasible Learned Acquisition Trajectories for Accelerated MRI". In: *MLMIR WS at MICCAI*. Oct. 2020, p. 3 (pages 3, 8, 59, 60, 65, 81, 84, 100, 108, 113).

[Wan+04] Z. Wang et al. "Image quality assessment: from error visibility to structural similarity". In: *IEEE TMI* 13.4 (2004), pp. 600–612 (page 44).

[Wan+19] R. Wang et al. "PBBFMM3D: a Parallel Black-Box Fast Multipole Method for Non-oscillatory Kernels". In: *CoRR* abs/1903.02153 (2019) (pages 2, 6, 69).

[Wan+21] G. Wang et al. "B-spline parameterized joint optimization of reconstruction and k-space trajectories (BJORK) for accelerated 2d MRI". In: *arXiv preprint arXiv:2101.11369* (2021) (pages 3, 8, 60, 65, 80, 81, 84, 108, 109, 113, 114, 119).

[Wan+22] G. Wang, J.-F. Nielsen, J. A. Fessler and D. C. Noll. *Stochastic Optimization of 3D Non-Cartesian Sampling Trajectory (SNOPY)*. 2022 (pages 113, 126).

[Wei+20] T. Weiss et al. "PILOT: Physics-informed learned optimal trajectories for accelerated MRI". In: *arXiv:1909.05773v4* (Aug. 2020) (pages 3, 8, 59, 60, 65, 81, 84, 100, 108, 109, 113, 114, 119).

[WF23] G. Wang and J. A. Fessler. "Efficient Approximation of Jacobian Matrices Involving a Non-Uniform Fast Fourier Transform (NUFFT)". In: *IEEE Transactions on Computational Imaging* (2023), pp. 1–12 (pages 3, 8, 60, 108, 109, 112).

[WSB03a] Z. Wang, E. Simoncelli and A. Bovik. "Multiscale structural similarity for image quality assessment". In: *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003*. Vol. 2. 2003, 1398–1402 Vol.2 (page 112).

[WSB03b] Z. Wang, E. P. Simoncelli and A. C. Bovik. "Multiscale structural similarity for image quality assessment". In: *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*. Vol. 2. Ieee. 2003, pp. 1398–1402 (page 44).

[Yar21] D. Yarotsky. "Universal approximations of invariant maps by neural networks". In: *Constructive Approximation* (2021), pp. 1–68 (page 51).

[Zac+19] Zaccharie Ramzi, P. Ciuciu, J.-L. Starck and J.-L. Starck Benchmarking. "Benchmarking proximal methods acceleration enhancements for CS-acquired MR image analysis reconstruction". In: *SPARS 2019 - Signal Processing with Adaptive Sparse Structured Representations Workshop*. 2019 (page 31).

[Zbo+18] J. Zbontar et al. "fastMRI: An open dataset and benchmarks for accelerated MRI". In: *arXiv:1811.08839* (2018) (pages 3, 8, 44, 52, 55, 56, 58, 65, 72, 101, 108, 110, 124, 143).

[Zha+15] Y. Zhang, Z. Dong, P. Phillips, S. Wang, G. Ji and J. Yang. "Exponential wavelet iterative shrinkage thresholding algorithm for compressed sensing magnetic resonance imaging". In: *Information Sciences* 322 (2015), pp. 115–132 (page 29).

[Zhe+15] H. Zheng, Z. Yang, W. Liu, J. Liang and Y. Li. "Improving deep neural networks using softplus units". In: *2015 International joint conference on neural networks (IJCNN)*. IEEE. 2015, pp. 1–4 (page 48).

[Zhu+18] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen and M. S. Rosen. "Image reconstruction by domain-transform manifold learning". In: *Nature* 555.7697 (2018), pp. 487–492 (page 53).

[ZKR21] M. V. W. Zibetti, F. Knoll and R. R. Regatte. "Alternating Learning Approach for Variational Networks and Undersampling Pattern in Parallel MRI Applications". In: *IEEE Transactions on Computational Imaging* 8 (2021), pp. 449–461 (page 57).