# Image Reconstruction from Undersampled Data with Application to Accelerated Magnetic Resonance Imaging

# Beeldreconstructie van ondergesamplede gegevens met toepassing op versnelde magnetische resonantiebeeldvorming

Marko Panić

Promotoren: Prof. dr. ir. A. Pižurica, Prof. dr. ir. D. Vukobratović
Onderzoeksgroep Group for Artificial Intelligence and Sparse Modelling - GAIM
Proefschrift ingediend tot het behalen van de graad van
Doctor in de Ingenieurswetenschappen

Vakgroep Telecommunicatie en Informatieverwerking
Voorzitter: Prof. dr. ir. J. Walraevens
Faculteit Ingenieurswetenschappen
Academiejaar 2018-2019


УНИВЕРЗИТЕТ У НОВОМ САДУ ● **ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА**
21000 НОВИ САД, Трг Доситеја Обрадовића 6

# КЉУЧНА ДОКУМЕНТАЦИЈСКА ИНФОРМАЦИЈА

| | | |
|---|---|---|
| Редни број, **РБР**: | | |
| Идентификациони број, **ИБР**: | | |
| Тип документације, **ТД**: | | Монографска документација |
| Тип записа, **ТЗ**: | | Текстуални штампани материјал |
| Врста рада, **ВР**: | | Докторска дисертација |
| Аутор, **АУ**: | | Марко Панић |
| Ментор, **МН**: | | Проф. др Дејан Вукобратовић и Проф. Др Александра Пижурица |
| Наслов рада, **НР**: | | Реконструкција сигнала из непотпуних мерења са применом у убрзању алгоритама за реконструкцију слике магнетне резонанце |
| Језик публикације, **ЈП**: | | Енглески |
| Језик извода, **ЈИ**: | | Енглески / српски |
| Земља публиковања, **ЗП**: | | Србија |
| Уже географско подручје, **УГП**: | | Аутономна Покрајина Војводина |
| Година, **ГО**: | | 2019 |
| Издавач, **ИЗ**: | | Ауторски репринт |
| Место и адреса, **МА**: | | Факултет техничких наука, Трг Доситеја Обрадовића 6, Нови Сад |
| Физички опис рада, **ФО**: (поглавља/страна/ цитата/табела/слика/графика/прилога) | | 7 поглавља / 186 стране / 234 цитата / 2 табеле / 48 слика / 43 графика |
| Научна област, **НО**: | | Електротехничко и рачунарско инжењерство |
| Научна дисциплина, **НД**: | | Статистичко моделовање ретке репрезентације слике |
| Предметна одредница/Кqучне речи, **ПО**: | | Реконструкција слике, ретка репрезентација, статистичко моделовање, оптимизација, магнетна резонанца |
| **УДК** | | Монографска документација |
| Чува се, **ЧУ**: | | Библиотека Факултета техничких наука, Универзитет у Новом Саду |
| Важна напомена, **ВН**: | | |
| Извод, **ИЗ**: | | У докторској дисертацији разматран је проблем реконструкције сигнала слике из непотпуних мерења који има директну примену у креирању слика магнетне резнонаце. Предмет истраживања је везан за предлог нових регуларизационих метода реконструкције коришћењем статистичких модела Марковљевог случајног поља и теорије ретке репрезентације сигнала. На основу предложеног модела који на веродостојан начин репрезентује статистику сигнала слике предложене су регуларизационе функције и креирана четири алгоритма за реконструкцију слике магнетне резонанце. |
| Датум прихватања теме, **ДП**: | | 27.12.2018. |
| Датум одбране, **ДО**: | | |
| Чланови комисије, **КО**: | Председник: | Проф. др Gert De Cooman |
| | Члан: | Проф. др Владимир Црнојевић |
| | Члан | Проф. др Jan Sijbers |
| | Члан | др Pim Pullens |
| | Члан: | др Љиљана Платиша |
| | Члан:,ментор | Проф. др Александра Пижурица |
| | Члан, ментор: | Проф. др Дејан Вукобратовић |

Потпис ментора

UNIVERSITY OF NOVI SAD ● **FACULTY OF TECHNICAL SCIENCES**
21000 NOVI SAD, Trg Dositeja Obradovića 6

# KEY WORDS DOCUMENTATION

| | | |
|---|---|---|
| Accession number, **ANO**: | | |
| Identification number, **INO**: | | |
| Document type, **DT**: | | Monograph documentation |
| Type of record, **TR**: | | Textual printed material |
| Contents code, **CC**: | | PhD thesis |
| Author, **AU**: | | Marko Panić |
| Mentor, **MN**: | | Dr Dejan Vukobratović Associate Professor and Dr Aleksandra Pižurica Professor |
| Title, **TI**: | | Image Reconstruction from Undersampled Data with Application to Accelerated Magnetic Resonance Imaging |
| Language of text, **LT**: | | English |
| Language of abstract, **LA**: | | English / Serbian |
| Country of publication, **CP**: | | Serbia |
| Locality of publication, **LP**: | | Autonomous Provence of Vojvodina |
| Publication year, **PY**: | | 2019 |
| Publisher, **PB**: | | Author's reprint |
| Publication place, **PP**: | | Faculty of Technical Sciences, Trg Dositeja Obradovica 6, Novi Sad |
| Physical description, **PD**: (chapters/pages/ref./tables/pictures/graphs/appendixes) | | 7 chapters/ 186 pages/ 234 ref/ 2 tables/ 48 pictures / 43 graphs |
| Scientific field, **SF**: | | Electrical and Computer Engineering |
| Scientific discipline, **SD**: | | Statistical modelling of sparse image representation |
| Subject/Key words, **S/KW**: | | Image reconstruction, sparse representation, statistical modelling, optimization, magnetic resonance |
| **UC** | | |
| Holding data, **HD**: | | Library of the Faculty of Technical Sciences, University of Novi Sad |
| Note, **N**: | | |
| Abstract, **AB**: | | In dissertation a problem of reconstruction of images from undersampled measurements is considered which has direct application in creation of magnetic resonance images. The topic of the research is proposition of new regularization based methods for image reconstruction which are based on statistical Markov random field models and theory of compressive sensing. With the proposed signal model which follows the statistics of images, a new regularization functions are defined and four methods for reconstruction of magnetic resonance images are derived. |
| Accepted by the Scientific Board on, **ASB**: | | 27.12.2018. |
| Defended on, **DE**: | | |
| Defended Board, **DB**: | President: | Gert De Cooman, Professor |
| | Member: | Dr Vladimir Crnojević, Professor |
| | Member: | Dr Pim Pullens, research associate |
| | Member: | Dr Ljiljana Platiša, research associate |
| | Member: | Dr Jan Sijbers, Professor |
| | Member, Mentor | Dr Aleksandra Pižurica, Professor |
| | Member, Mentor: | Dr Dejan Vukobratović, Associate Professor |

Menthor's sign

*Every living being is an engine geared to the wheelwork of the universe. Though seemingly affected only by its immediate surrounding, the sphere of external influence extends to infinite distance.*
—Nikola Tesla

**Members of the jury**

Prof. dr. ir. Vladimir Crnojević (University of Novi Sad)
Prof. dr. ir. Gert De Cooman (Ghent University, chairman)
Prof. dr. ir. Aleksandra Pižurica (Ghent University, supervisor)
Dr. Ljiljana Platiša (Ghent University)
Dr. ir. Pim Pullens (Ghent University Hospital, secretary)
Prof. dr. ir. Jan Sijbers (University of Antwerp)
Prof. dr. ir. Dejan Vukobratović (University of Novi Sad, supervisor)

**Affiliations**

Department of Telecommunications and Information Processing (TELIN)
Faculty of Engineering
Ghent University

Sint-Pietersnieuwstraat 41
B-9000 Ghent
Belgium


# Acknowledgements

First and foremost, I would like to express my sincere gratitude to my dear supervisors Prof. dr. ir. Aleksandra Pižurica and Prof. dr. ir. Dejan Vukobrtović for their selfless guidance. They taught me how to understand the research problem from a broader perspective and how to be persistent and consistent in pursuing the solution. Without their help, this work would have never been carried out. Besides them I would like to give thanks to my Prof. dr. ir. Vladimir Crnojević for his honest advice and great understanding. They helped me to clarify my goals and research directions and showed me how to move patiently and motivated towards them. I'm immensely grateful to them for constant support, encouragement, and motivation.

I would also like to thank all my BioSense colleagues for creating a nice atmosphere that makes work pleasant. Especially, I am thankful to my friends and colleagues from the research group in BioSense: Vladan Minić, Predrag Lugonja, Branko Šikoparija, Sanja Brdar and Oskar Marko. My colleagues Jan Aelterman and Ljubomir Jovanov from Ghent University help me a lot in my scientific research with interesting discussions during the afternoon coffee/tea breaks during my visits to Ghent. Dear colleagues, thank you very much.

I owe most of all the special gratitude to my mother Dragica and father Neven for their love, trust and sacrifice that they accepted so I can have a better opportunity to succeed. I would like to extend my thanks to my brother Nikola for his support and understanding. Finally, I would like to thank my wife Aleksandra for unlimited patience and unconditional love.

*Marko Panić*

# Prošireni izvod na srpskom jeziku

Putem magnetne rezonance (MR) moguće je dobiti slike visoke rezolucije dobijene bez korišćenja ijonizućeg zračenja, što je jedan od ključnih faktora u modernoj medicini prilikom proglašavanja dijagnoza [Kuperman 00, Liang 00]. MRI skeneri koriste jako statičko magnetno polje i radio-frekventne (RF) talase da bi generisali sliku ljudske anatomije [Rabi 38, Bloch 46, Purcell 46, Lauterbur 73]. U poređenju sa drugim medicinskim sistemima za snimanje, slike magnetne rezonance imaju nekoliko prednosti. Njihova glavna prednost je superiorna kontrastna rezolucija koja omogućava jasnije uočavanje razlika između tkiva, kao na primer između bele i sive materije u mozgu ili između tkiva jetre i slezine. Prilikom definisanja protokola snimanja magnetnom rezonancom postoji dosta parametara čije pravilno podešavanje omogućava dobijanje sveobuhvatnih informacija o snimanom delu ljudske anatomije bilo da je u pitanju stacionarno tkivo ili nestacionarno kao što je krv. Ova fleksibilnost u izboru parametara veoma je važna u procesima optimizacije rezolucije kontrasta radi bolje viualizacije različitih stanja tkiva i optimizacije prostorne rezolucije slike spram vremena skeniranja. Dodatna prednost MR je i sposobnost dobijanja slika transverzalnih, sagitalnih i koronalnih ravni (preseka) kao i slika kosih projekcija trodimenzionalnog objekta koji je skeniran. Usled ove sposobnosti MR skenera dobija se veliki skup podataka tokom jedne sekvence snimanja. Dobijena količina podataka pomaže u rekonstrukciji bilo koje anatomske ravni sa visokom osetljivošću, time omogućavajući detekciju suptilnih razlika u anatomiji.

Ipak proces skeniranja MR uređajem je dugotrajan i može trajati više od jednog sata u zavisnosti od vrste pregleda. Tokom snimanja MR skener proizvodi neprijatne zvukove izazvane pomeranjem magneta pa se preporučuje da pacijent nosi slušalice zbog jačine zvukova. Neke vrste pregleda zahtevaju od pacijenta da zadržavaju dah tokom skeniranja. Postoje takođe situacije u kojima pojedinci doživljavaju vizuelne senzacije bljeskova svetlosti na mrežnjači oka kada se nalaze u jakom

magnetnom polju. Ove senzacije koje se nazivaju magnetofosfeni prvi je primetio francuski fiziolog Jaques-Arséne d'Arsonval 1896. Senzacije su uzrokovane indukovanim električnim strujama u mrežnjači oka kada se pacijent kreće kroz magnetno polje ili kada je nepokretan u promenljivom magnetnom polju. Uski, zatvorni prostor u MR skeneru u kome se pacijent nalazi tokom skeniranja može izazvati anksioznost kod pacijenta. U takvoj situaciji medicinsko osoblje može dati lekove pacijentu pre snimanja kako bi postupak bio podnošljiviji. Smanjenje vremena skeniranja i stoga izlaganje neprijatnim i neudobnim okolnostima tokom skeniranja je primarni cilj kojim se teži uz postizanje snimaka dobrog kvaliteta i održavanja svih prednosti koje MR donosi u odnosu na druge medicinske sisteme snimanja.

Radi boljeg razumevanja kako se može postići ubrzanje skeniranja MR uređajem, uvodimo objašnjenje kakav tip merenja MR uređaj prikuplja tokom skeniranja. U pitanju su prostorne frekvencije slike koje se nalaze u domenu prostornih frekvencija nazvanom $k$-prostor. Stoga merenja prikupljena tokom skeniranja predstavljaju drugačiju reprezentaciju slike u tkz. prostorno-frekventnom domenu za razliku od standardne reprezentacije slike korišćene u digitalnim kamerama. Konačna slika je dobijena nakon iterativne procedure rekonstrukcije koja uključuje inverznu Furijeovu transformaciju radi povezivanja prostornih frekvencijskih komponenti dobijenih merenjem sa prostornim informacijama u slici tj. vrednostima piksela. Količina prikupljenih merenja sa predefinisanim putanjama (trajektorijama) odabiranja u $k$-prostoru utiče na kvalitet rekonstruisane slike. Fizička ograničenja hardvera MR uređaja su najveći razlog sporog prikupljanja merenja i ograničavajući faktor u dizajnu trajektorija odabiranja. Stoga, postoje dva načina da se ubrza proces akvizicije: i) dizajn hardvera koji omogućava istovremenu akviziciju sa više kalema i ii) korišćenje nelinearnih algoritama za rekonstrukciju slike iz nepotpunih merenja.

U ovoj tezi razmotrićemo drugi pristup koji podrazumeva dizajn nelinearnih algoritama za rekonstrukciju slike iz nepotpunih merenja dobijenih korišćenjem različito dizajniranih trajektorija odabiranja u $k$-prostoru. Nepotpuna merenja dovode do loše postavljenog problema rekonstrukcije pošto Shannon/Nyquist period odabiranja nije ispunjen [Kotelnikov 33, Nyquist 28, Shannon 49, Whittaker 15]. Ovakvi problemi se obično formulišu kroz neodredjen sistem linearnih jednačina koji ima bezbroj mogućih rešenja. Korišćenjem teorije retke reprezentacije (compressed sensing CS) i statističkog modela signala na bazi Markovljevog slučajnog polja, biće predloženo nekoliko nelinearnih algoritama

za rekonstrukciju slike magnetne rezonance. U nastavku je dat opis formiranja slike iz merenja i formulisana postavka problema rekonstrukcije slike.

## Formiranje slike magnetne rezonance

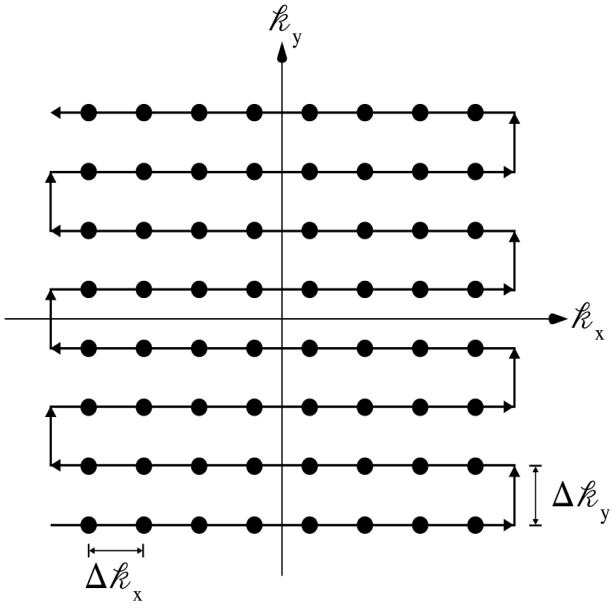
Detekcija signala (merenja) magnetne rezonance je uslovljena promenom magnetnog fluksa u jednom ili više kalemova MR skenera. Na osnovu Faradejevog zakona indukcije, napon (signal) indukovan u kalemu je linearno zavisan od vremenske promene magnetnog fluksa. Promena fluksa je posledica precesije vektora magnetizacije $\mathbf{M}$ oko eksternog statičkog magnetnog polja $\mathbf{B}_0$, koji na makroskopskom nivou predstavlja ponašanje spinova protona u zapremini skeniranog objekta. Da bi se postigla precesija vektora $\mathbf{M}$, tj. da bi se stimulisao sistem spinova protona u pomatranoj zapremini, uvodi se vremenski promenljivo magnetno polje $\mathbf{B}_1$ u formi radio-frekventnog (RF) talasa sa ugaonom frekvencijom $\omega_{RF}$ u kratkom vremenskom periodu. Frekvencija rotacije/oscilovanja RF talasa $\mathbf{B}_1$ je jednaka Larmorovoj frekvenciji $\omega_{RF} = \omega_L$, time ispunjavajući rezonantni uslov neophodan za detekciju signala u MR uređaju, dok je pravac delovanja magnetnog polja $\mathbf{B}_1$ upravan na pravac delovanja statičkog magnetnog polja $\mathbf{B}_0$. Uvođenjem RF talasa dolazi do pomeraja vektora magnetizacije $\mathbf{M}$ iz $z$-pravca u $xy$-ravan, čime se postiže indukovanje elektormotorne sile u prijemnom kalemu. Preciznije, indukovani napon u kalemu odnosno MR signal $s(t)$ je direktno povezan sa promenom magnitude transverzalne komponente $|\mathbf{M}_{xy}|$ vektora magnetizacije i zavisan je od senzitivnosti prijemnog kalema $\mathbf{B}_{xy}^R$ u $xy$-ravni, preko sledeće relacije:

$$s(t) = \int_\Omega B_{xy}^R(\mathbf{r}) |M_{xy}(\mathbf{r}, t)| d\mathbf{r} \tag{1}$$

gde je $\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ a $\Omega$ označava zapreminu koja je obuhvaćena kalemom.

Određivanje prostornog doprinosa vektora magnetizacije u formiranju MR signala, a samim tim i kreiranje MR slike, zahteva prostorno selektivnu ekscitaciju od strane promenljivog magnetnog polja u planarnom regionu trodimenzionalnog objekta. Ovo se postiže uvođenjem gradijentnog magnetnog polja $\mathbf{B_G} = B_{G_x}\mathbf{i} + B_{G_y}\mathbf{j} + B_{G_z}\mathbf{k}$ duž $z$-pravca, čime se menja frekvencija sistema spinova u tački prostora $\mathbf{r}$ na sledeći način

$$\omega(\mathbf{r}) = \omega_L + \gamma B_{G_z}(\mathbf{r}). \tag{2}$$




**Slika 1:** Pokrivanje $k$-prostora trajektorijom. Svaka tačka reprezentuje pozicuju odabiranja. Linije koje povezuju tačke pokazuju pravac očitavanja tj. generisanja merenja dok strelice ukazuju na hronološki red očitavanja.

Sa izolovanom zapreminom korišćenjem selektivne ekscitacije moguće je formirati sliku planarnog dela trodimenzionalnog objekta manipulacijom faze vektora magnetizacije $M_{xy}(\mathbf{r},t) = |M_{xy}(\mathbf{r},t)|e^{-i\omega_{RF}t}$ [Wright 97]. Longitudinalna komponenta gradijentnog magnetnog polja $B_{G_z}$ varira linearno sa $x, y$ i $z$-pravcem:

$$B_{G_z}(\mathbf{r}) = G_x x + G_y y + G_z z = \langle \mathbf{G}, \mathbf{r} \rangle \tag{3}$$

gde je $\mathbf{G} = G_x\mathbf{i} + G_y\mathbf{j} + G_z\mathbf{k}$ gradijentni vektor. Ovim se menja faza transverzalne komponente vektora magnetizacije na sledeći način $M_{xy}(\mathbf{r},t) = |M_{xy}(\mathbf{r},t)|e^{-i\omega_{RF}t}e^{-i\gamma\langle \mathbf{G},\mathbf{r}\rangle t}$. Ako se gradijentno polje menja sa vremenom, MR signal $s(t)$ se predstavlja preko sledeće relacije:

$$s(t) = \int_{\Omega} B_{xy}^{R}(\mathbf{r})|M_{xy}(\mathbf{r},t)|e^{-i\gamma\int_0^t \langle \mathbf{G}(t'),\mathbf{r}\rangle dt'} d\mathbf{r}. \tag{4}$$

gde MR signal pruža informaciju o prostorno-frekventnim komponentama slike umesto direktno o informacijama o vrednostima piksela na

određenim pozicijama. Stoga, MR signal je detektovan u prostorno-frekventnom domenu slike, $\mathbb{k}$-prostoru, čija pokrivenost zavisi od gradijentnog polja. Ako označimo trajektoriju u $\mathbb{k}$-prostoru kao $\boldsymbol{\kappa}(t) = \mathbb{k}_x(t)\mathbf{i} + \mathbb{k}_y(t)\mathbf{j} + \mathbb{k}_z(t)\mathbf{k}$ i uvedemo sledeću relaciju:

$$\boldsymbol{\kappa}(t) = \frac{\gamma}{2\pi} \int_0^t \mathbf{G}(t')dt' \tag{5}$$

onda izraz za $s(t)$ se može reformulistati na sledeći način:

$$s(t) = \int_\Omega B_{xy}^R(\mathbf{r}) |M_{xy}(\mathbf{r}, t)| e^{-2\pi i \langle \boldsymbol{\kappa}(t), \mathbf{r} \rangle} d\mathbf{r}. \tag{6}$$

Usled selektivnosti ekscitacije, MR signal je generisan samo sistemom spinova u $xy$-ravni, te se stoga izraz za $s(t)$ može pojednostaviti

$$s(t) = \int_x \int_y B_{xy}^R(x, y) |M_{xy}(x, y, t)| e^{-2\pi i (\mathbb{k}_x(t)x + \mathbb{k}_y(t)y)} dxdy \tag{7}$$

gde figurišu samo $x$ i $y$ komponente vektora magnetizacije i odgovarajuće komponente trajektorija u $\mathbb{k}$-prostoru. Za fiksan trenutak u vremenu $t_{acq}$, $s(t_{acq})$ je dvodimenzionalna Furijeova transformacija slike $\rho(x, y) = B_{xy}^R(x, y)|M_{xy}(x, y)|$ koja je izmerena u prostornim frekvencijama $(\mathbb{k}_x(t_{\text{acq}}), \mathbb{k}_y(t_{\text{acq}}))$, uz unapred usvojenu pretpostavku da je $|M_{xy}(x, y)|$ konstantno tokom vremena akvizicije $t \in [0, t_{\text{acq}}]$ [Wright 97]. Na slici 1 je prikazana trajektorija u $\mathbb{k}$-prostoru sa tačkama koje označavaju diskretne odbirke u trenutku akvizicije $t_{\text{acq}}$.

Notacije za merenja $s(t_p)$ sa $\forall p \in \mathbb{Z}$ se mogu pojednostaviti sa sledećim načinom indeksiranja $s(\boldsymbol{\kappa}_p)$ gde $\boldsymbol{\kappa}_p$ predstavlja tačke u $\mathbb{k}$-prostoru gde je trajektorija izvršila odabiranje u vremenskom trenutku $t = t_p$, $p \in \mathbb{Z}$. Na ovaj način izraz (6) postaje:

$$s(\boldsymbol{\kappa}_p) = \int_{\Omega_{xy}} \boldsymbol{\rho}(\mathbf{r}_{xy}) e^{-2\pi i \langle \boldsymbol{\kappa}_p, \mathbf{r}_{xy} \rangle} d\mathbf{r}_{xy}, \quad \forall p \in \mathbb{Z} \tag{8}$$

gde $\mathbf{r}_{xy} = r_x\mathbf{i} + r_y\mathbf{j}$ predstavlja vektor koordinata u dvodimenzionalnom potprostoru $\Omega_{xy}$ i gde je $\boldsymbol{\rho}(\mathbf{r}_{xy}) = B_{xy}^R(\mathbf{r}_{xy})|M_{xy}(\mathbf{r}_{xy})|$ naša kontinualna slika koju želimo da dobijemo. Koristeći pristup iz [Fessler 10] možemo aproksimirati fizičku gustinu objekta koju skeniramo $\boldsymbol{\rho}(\mathbf{r}_{xy})$ sa sledećim razvojem u red:

$$\boldsymbol{\rho}(\mathbf{r}_{xy}) = \sum_{n=1}^{N} x_n \mathbf{b}(\mathbf{r}_{xy} - \mathbf{r}_n) \tag{9}$$

gde $\mathbf{b}(\cdot)$ označava baznu funkiju, $\mathbf{r}_n$ je centar $n$-te translirane bazne funkcije a $N$ je broj parametara. Ovde su diskretni odbirci kontinualne prostorne funkcije $\boldsymbol{\rho}(\mathbf{r}_{xy})$ predstavljeni sa $x_n$. Uvrštavanjem poslednjeg izraza za $\boldsymbol{\rho}(\mathbf{r}_{xy})$ u izraz za $s(\boldsymbol{\kappa}_p)$ dobijamo sledeću aproksimaciju:

$$s(\boldsymbol{\kappa}_p) = \sum_{n=1}^{N} a_{pn} x_n, \quad \forall p \in \mathbb{Z} \tag{10}$$

sa

$$a_{pn} = \int_{\Omega_{xy}} \mathbf{b}(\mathbf{r}_{xy} - \mathbf{r}_n) e^{-2\pi i \langle \boldsymbol{\kappa}_p, \mathbf{r}_{xy} \rangle} d\mathbf{r}_{xy}, \quad \forall p \in \mathbb{Z}, \quad n = 1, ..., N. \tag{11}$$

Sa osobinom da su bazne funkcije u praksi izrazito lokalizovane, možemo usvojiti sledeću aproksimaciju:

$$a_{pn} \approx e^{-2\pi i \langle \boldsymbol{\kappa}_p, \mathbf{r}_n \rangle} \quad \forall p \in \mathbb{Z}, \quad n = 1, ..., N. \tag{12}$$

što dalje implicira da je podintegralna funkcija u izrazu za $a_{pn}$ različita od nule samo kada je zadovoljena sledeća jednakost $\mathbf{r}_{xy} = \mathbf{r}_n$.

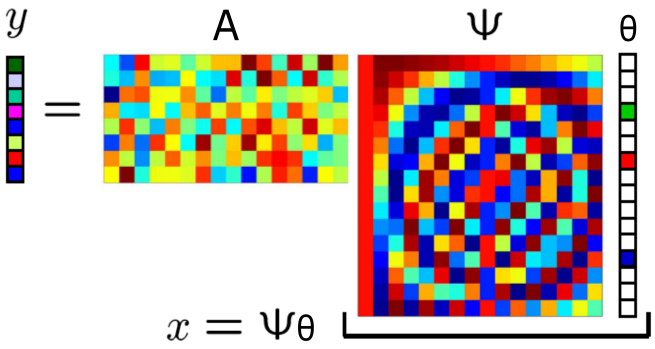
Zbog konzistentnosti u notaciji sa literaturom uvodimo vektor $\mathbf{x} = (x_1, x_2, x_3, ..., x_N)$ koji predstavlja, u rasterizovanoj formi, diskretnu sliku kontinualnog objekta $\boldsymbol{\rho}(\mathbf{r}_{xy})$. Znajući da je broj tačaka odabiranja u $\boldsymbol{\kappa}$-prostoru u praksi konačan i jednak nekom broju $M$, i.e. $p \in [0, M]$, možemo definisati $\mathbf{y} = (s(\boldsymbol{\kappa}_1), s(\boldsymbol{\kappa}_2), s(\boldsymbol{\kappa}_3), ..., s(\boldsymbol{\kappa}_M))^T$ kao vektor merenja a $\mathbf{A} = \{a_{m,n}\} \in \mathbb{C}^{M \times N}$ kao matricu Furijeove transformacije sa elementima:

$$a_{m,n} = e^{-2\pi i \langle \boldsymbol{\kappa}_m, \mathbf{r}_n \rangle}. \tag{13}$$

Na osnovu izloženog, problem rekonstrukcije MR slike bez prisustva šuma u merenjima, se može predstaviti kao algebarski linearno inverzan problem u sledećoj matričnoj formi:

$$\mathbf{y} = \mathbf{A}\mathbf{x}. \tag{14}$$

U ovoj tezi usvajamo model akvizicije MR merenja iz prethodne jednačine sa proširenjem koje obuhvata modelovanjem prisustva šuma u merenjima. Razmatrićemo optimizacione metode za rekonstrukciju slike uz uslov da je broj vrsta matrice $\mathbf{A}$, tj. broj merenja, umnogome manji od broja kolona matrice $\mathbf{A}$ tj. broja piksela u slici što dovodi do neodređenog sistema linearnih jednačina odnosno loše postavljenog inverznog linearnog problema. Da bismo rešili postavljen problem na ovakav način potpomažemo se teorijom retke reprezentacije koju uvodimo u nastavku.




**Slika 2:** Model procesa merenja MR uređajem predstavljen sa matricom merenja $\mathbf{A}$ matricom retke reprezentacije signala $\mathbf{\Psi}$. Vektor koeficijenata $\boldsymbol{\theta}$ je redak sa $K = 3$ nenulta koeficijenta.

## Retka reprezentacija slike

Signal u visokodimenzionalnom prostoru, kao što je slika, može biti predstavljen vektorom $\mathbf{x} \in \mathbb{R}^N$ i reprezentovan u bazi ili rečniku $\mathbf{\Psi}$ na sledeći način:

$$\mathbf{x} = \mathbf{\Psi}\boldsymbol{\theta} \tag{15}$$

gde $\boldsymbol{\theta}$ predstavljaju koeficijente nove reprezentacije signala. Ako je reprezentacija signala retka u izabranoj bazi ili rečniku $\mathbf{\Psi} \in \mathbb{R}^{D \times N}$, onda samo $K \ll N$ komponenti vektora koeficijenata $\boldsymbol{\theta}$ su nenulti tj. $\|\boldsymbol{\theta}\|_0 = K$. U usvojenom modelu za akviziciju slike, retka reprezentacija signala slike izražena preko $\boldsymbol{\theta}$ nije direktno merena sa MR uredjajem. Umesto toga merene su prostorne frekvencije slike $\mathbf{x}$ izražene preko vektora $\mathbf{y}$ na sledeći način:

$$\mathbf{y} = \mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{\Psi}\boldsymbol{\theta} \tag{16}$$

gde je matrica merenja $\mathbf{A}$ dimenzije $M \times N$. Matrica $\mathbf{A}$ je operator redukcije dimenzionalnosti pošto projektuje iz $\mathbb{R}^N$, gde je $N$ u opštem slučaju veliko, na $\mathbb{R}^M$, $M \ll N$. Slika. 2 ilustruje proces prikupljanja merenja i relaciju između retke reprezentacije signala i dobijenih merenja.

Da bi rekonstrukcija iz nepotpunih merenja $M \ll N$ bila moguća, teorija retke reprezentacije definiše uslove koje signal $\mathbf{x}$ i matrica $\mathbf{A}$ mora da zadovoljavaju. Prvi uslov je postojanje retke reprezentacije signala $\mathbf{x}$ u izabranom rečniku ili bazi. U slučaju MR slika, ovaj uslov je delimično zadovoljen, preciznije rečeno MR slike se mogu veoma dobro

aproksimirati retkom reprezentacijom jer njihovi koeficijenti tj. komponente vektora $\boldsymbol{\theta}$ sortirani u opadajućem poretku njihove magnitude $|\theta_1| \geq |\theta_2| \geq |\theta_3| \geq ... \geq |\theta_N|$ imaju sledeću karakteristiku:

$$|\theta_i| \leq Ri^{-\frac{1}{r}}, \quad i = 1, 2...., N \tag{17}$$

gde su $R > 0$ i $r > 0$ konstante [Cevher 10a]. Zbog ove osobine brzog opadanja magnitude koeficijenata, slike se mogu veoma precizno aproksimirati sa $K \ll N$ koeficijenata vektora $\boldsymbol{\theta}$. Sledeći uslov je vezan za matricu merenja $\mathbf{A}$ i naziva se karakteristika $K$-ograničene izometrije tj. K-RIP od *K-restricted isometry property*. Ona se definiše korišćenjem konstante $\epsilon_K < 1$, i kaže se da matrica $\mathbf{A}$ zadovoljava K-RIP karakteristiku ako za sve $K$-retke signale $\mathbf{x} \in \sum_K$ iz potprostora $\sum_K$ važi sledeće:

$$(1 - \epsilon_K)\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1 + \epsilon_K)\|\mathbf{x}\|_2^2. \tag{18}$$

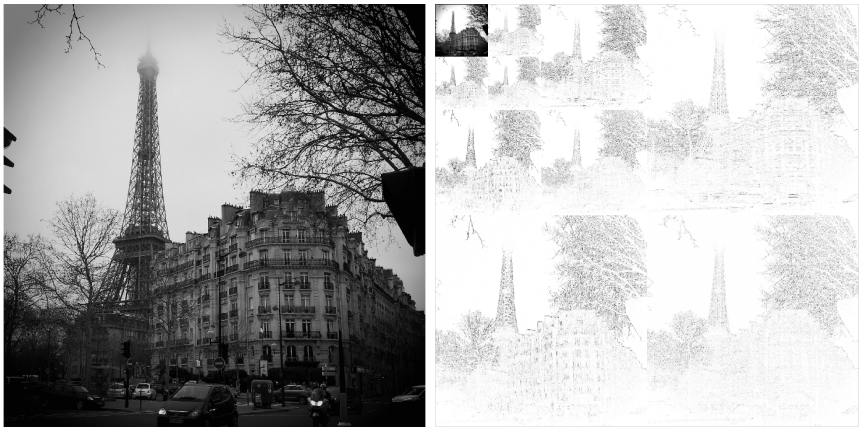
Ovo znači da matrica merenja $\mathbf{A}$ postiže stabilnu projekciju za prizvoljne vektore retke reprezentacije ne menjajući njihovu magnitudu. Uz zadovoljen K-RIP uslov za matricu $\mathbf{A}$ i postojanje retke reprezentacije signala slike, rekonstrukcija signala iz nepotpunih merenja na osnovu teorije retke reprezentacije je moguća [Cheng 16, Candes 08].

Retka reprezentacija signala vezana je za izbor baze ili rečnika $\boldsymbol{\Psi}$ tj. za izbor transformacije signala iz njegovog originalnog domena. Vejvlet analiza se zasniva na dekompoziciji signala na frekvencijske komponente zavisno od skale. U odnosu na Furijeovu analizu i njenu verziju sa korišćenjem prozora u prostoru (vremenu), vejvlet analiza pruža bolju lokalizaciju prostornih (vremenskih) frekvencijskih komponenti i reprezentaciju lokalnih promena signala i njegovih singularnosti. Detaljna matematička formulacija vejvlet analize se može naći u sledećim radovima [Daubechies 92, Chui 16, Mallat 99, Cohen 96]. Kod vejvlet transformaciji baza za Hilbertov prostor $\mathbf{L}^2(\mathbb{R}^2)$ za funkcije ograničene energije $f$ je konstruisana korišćenjem vejvlet funkcija:

$$\left\{\psi_{j,k}(x) = \frac{1}{\sqrt{2^j}}\psi\left(\frac{x - 2^j k}{2^j}\right)\right\}_{j\in\mathbb{Z}, k\in\mathbb{Z}^2}, \tag{19}$$

gde $x = (m, n) \in \mathbb{R}^2$ i $\psi_{j,k}(x)$ su funkcije konstruisane od vejlvet funkcije $\psi(x)$ nazvane majka vejvlet korišćenjem operacija skaliranja (dilatacije) definisanim sa $2^j$ i translacija realizovanim sa $2^j k$, $k = (u, v) \in \mathbb{Z}^2$. Korišćenjem vejvlet baze svaki signal (slika) $f(x)$ konačne energije može biti predstavljen na sledeći način:

$$f(x) = \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}^2} w_{j,k}\psi_{j,k}(x) \tag{20}$$



**Slika 3: Levo** Slika i njena **Desno** dvodimenzionalna diskretna vejvlet transformacija sa Daubechies (db2) vejvlet funkcijama. Crni pikseli označavaju vejlvet koeficijente sa visokom magnitudom.

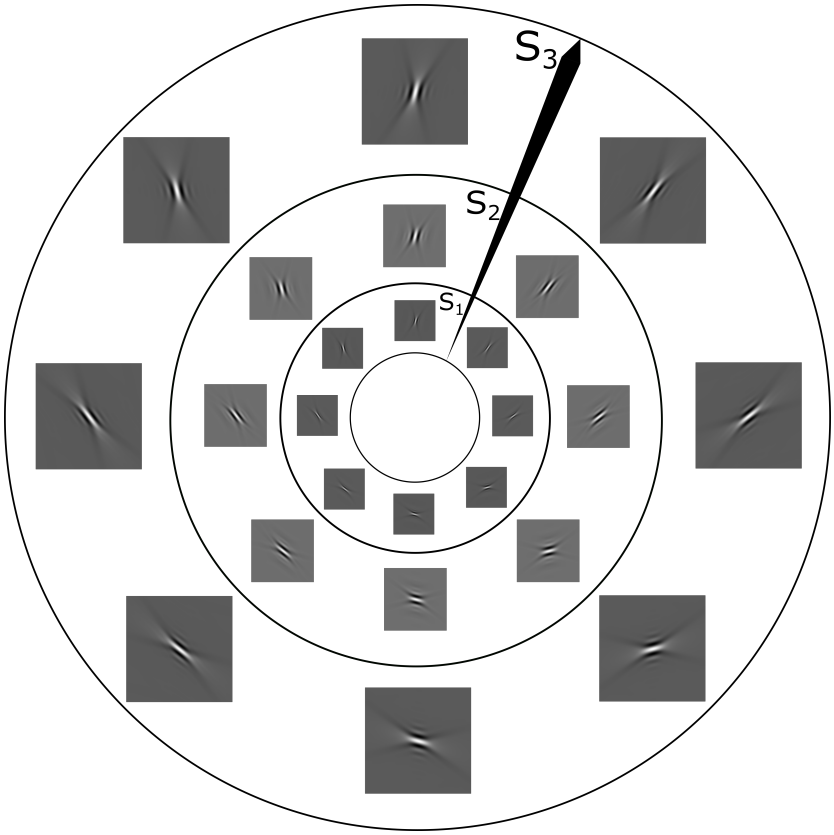
gde su $w_{j,k}$ vejvlet koeficijenti dobijeni unutrašnjim proizvodom $f(x)$ sa dualnim bazičnim funkcijama $\widetilde{\psi}_{j,k}$:

$$w_{j,k} = \langle f, \widetilde{\psi}_{j,k} \rangle = \sum_{m=1}^{M} \sum_{n=1}^{N} f(m,n) \widetilde{\psi}_{j,k}(m,n) \tag{21}$$

gde je $\widetilde{\psi}_{j,k}(m,n) = 0$ izvan oblasti definisanosti funkcije. Ako su dualne funkcije $\widetilde{\psi}_g$, ortogonalne sa odgovarajućim funkcijama $\psi_h$ u skupu za reprezentaciju, gde $g = (j,k)$ i $h = (i,z)$, onda $\widetilde{\psi}_g$ je biortogonalna sa $\psi_h$. Ovo znači da $\langle \psi_h, \widetilde{\psi}_g \rangle = \delta(h-g)$ gde $\delta(i)$ je Kronekerova delta funkcija. Vejvlet koeficijenti$\langle f, \widetilde{\psi}_{j,k} \rangle$ imaju malu amplitudu $|\langle f, \widetilde{\psi}_{j,k} \rangle|$ ako je $f$ regularno u oblasti definisanosti funkcije $\widetilde{\psi}_{j,k}$. Vejvlet koeficijenti imaju visoku vrednost amplitude na pozicijama ivica u slici što je prikazano na slici 3.

U ovoj tezi fokusiraćemo se na nedecimiranu vejvlet i širlet transformaciju za retku reprezentaciju signala. Širlet transformacija je nastala kao nadogradnja vejvlet transformacije sa boljim karakteristikama u predstavljanju geometrije u višedimenzionalnim podacima tj. verodostojnije reprezentacije ivica u slici. Širlet funkcije $\psi_{j,l,k}$, gde promenljive $j, l \in \mathbb{Z}, k \in \mathbb{Z}^2$ označavaju redom skalu, orijentaciju i prostornu poziciju, su formirane dilatacijom, iskrivljenjem i translacijom širlet funkcije $\psi \in \mathbf{L}^2(\mathbb{R}^2)$ preko sledeće relacije:

$$\psi_{j,l,k}(x) = |\det \mathbf{A}|^{j/2} \psi(\mathbf{B}^l \mathbf{A}^j x - k) \tag{22}$$




**Slika 4:** Širlet funkcije korišćenje za nedecimiranu 2D širlet transformaciju u 3 skale i 8 orijentacija po skali. Od centra ka periferiji su poslagane širlet funkcije po skalama (od najfinije ka najgrubljoj).

gde su $\mathbf{A}, \mathbf{B}$ invertibilne matrice $2 \times 2$ sa $|\det \mathbf{B}| = 1$. Ovim pristupom, matrice dilatacije $\mathbf{A}^j$ se odnose na transformaciju skale, dok matrice $\mathbf{B}^l$ se vezuju za geometrijske transformacije koje zadržavaju površinu, kao što su rotacija i iskrivljenje. U ovoj tezi normalizaciona konstanta $|\det \mathbf{A}|^{j/2}$ je izabrana tako da je norma $\|\psi\|_2 = \|\psi_{j,l,k}\|_2$ za sve vrednosti $j, l, k$ i korišćene su sledeće matrice transformacije:

$$\mathbf{A} = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}. \tag{23}$$

Širlet transformacija je optimalna u reprezentaciji dvodimenzionalnih funkcija koje su dvostruko diferencijabilne [Guo 07]. Može biti pokazano da greška aproksimacije povezana sa rekonstrukcijom iz $N$ najvećih širlet

koeficijenata $f_N^S$ zadovoljava sledeću nejednakost:

$$\|f - f_N^S\|_2^2 \leq BN^{-2}(\log N)^3, \quad N \to \infty \tag{24}$$

gde je $B$ konstanta. Ova osobina se često naziva optimalna retka reprezentacija [Guo 07]. Na slici 4 su prikazane širlet funkcije.

Spor proces akvizicije MR slike se može ubrzati korišćenjem teorije retke reprezentacije u algoritmima rekonstrukcije iz nepotpunih merenja, što je pokazano u radovima Lustiga i autora [Lustig 07, Lustig 08] kao i mnogim drugim studijama [Starck 05, Ma 08, Blumensath 09, Yang 10, Aelterman 11, Chen 12, Huang 11b, Rajani 12, Adcock 13, Chen 14].

U tezi je usvojen proširen model akvizicije merenja u $\not{k}$-prostoru $\mathbf{y} \in \mathbb{C}^M$ idealne slike $\mathbf{x} \in \mathbb{C}^N$

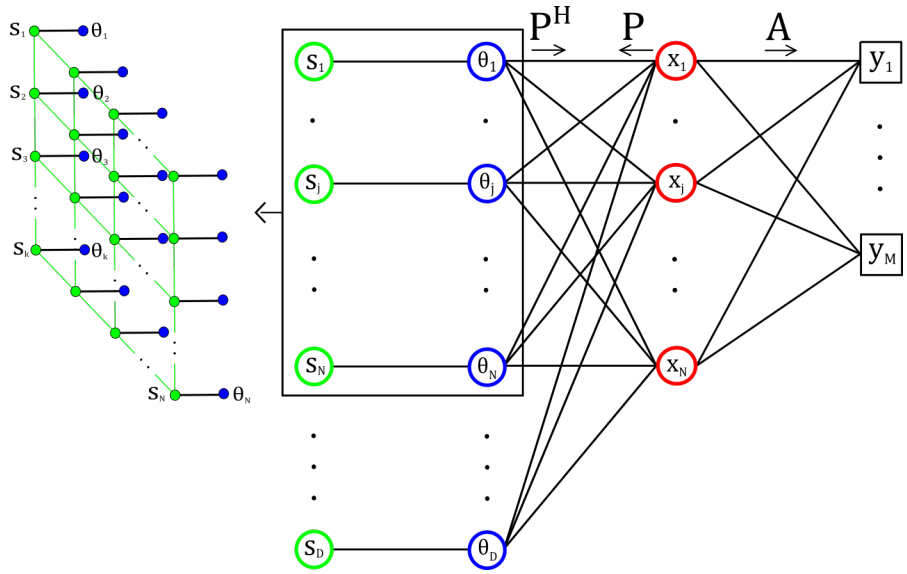
$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} \tag{25}$$

gde $M \ll N$ i $\mathbf{n} \in \mathbb{C}^M$ je beli Gausov šum, a $\mathbf{A} \in \mathbb{C}^{M \times N}$ označava Furijeov operator odabiranja [Lustig 07, Lustig 08]. Estimacija slike $\mathbf{x}$ iz merenja $\mathbf{y}$ je loše postavljen linearno inverzni problem, jer je matrica $\mathbf{A}$ singularna i/ili loše uslovljena. Pošto ne postoji jedinstveno rešenje za usvojen model neodređenog sistem linearnih jednačina, dodatna informacija o slici $\mathbf{x}$ je uvedena kroz formu regularizatora da stabilizuje i vodi pretragu prema očekivanom rešenju. MR slike su kompresibilne , tj mogu se dobro aproksimirati sa retkom reprezentacijom u odgovarajućim transformacionim domenima korišćenjem vejvlet ili širlet transformacije. Koeficijente u transformacionom domenu označavamo sa $\boldsymbol{\theta} = \mathbf{P}\mathbf{x} \in \mathbb{C}^D$ gde je $\mathbf{P} \in \mathbb{C}^{D \times N}$ matrica vejvlet, širlet ili neke druge transformacije. U nastavku ćemo razmatrati sledeći optimizacioni problem za estimaciju slike $\mathbf{x}$ iz merenja $\mathbf{y}$:

$$\min_{\mathbf{x} \in \mathbb{C}^N} \phi(\mathbf{P}\mathbf{x}) \quad \text{subject to} \quad \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \leq \epsilon \tag{26}$$

gde je $\phi : \mathbb{C}^N \mapsto \mathbb{R} \cup \{-\infty, +\infty\}$ regularizaciona funkcija a $\epsilon \geq 0$ je parametar povezan sa varijansom šuma koji kontroliše verodostojnost rekonstrukcije merenih vrednosti. Rešenje uvedenog optimizacionog problema je definisano kroz izvedene Moreau proksimalne mape funkcija regularizacije $\phi$ [Combettes 05]

$$\boldsymbol{\Psi}_\phi(\mathbf{u}; \mu) = \underset{\mathbf{x} \in \mathbb{C}^N}{\operatorname{argmin}} \, \phi(\mathbf{x}) + \frac{\mu}{2}\|\mathbf{x} - \mathbf{u}\|_2^2 \tag{27}$$




**Slika 5:** Grafička reprezentacija promenljivih, operatora i njihova veza u usvojenom modelu akvizicije. **Levo**: Skrivene labele $s_i$, pridružene koeficijentima $\theta_i$. Veze između suseda $s_i$ indiciraju njihovu statističku zavisnost, predstavljenu preko modela Markovljevog slučajnog polja. **Desno**: Grafički model koji prikazuje sve uključene promenljive, merenja i operatore i postavci problema.

U ovoj tezi fokusiraćemo se na statističko modelovanje retke reprezentacije MR slike odnosno vektora koeficijenata $\boldsymbol{\theta}$ korišćenjem Markovljevog slučajnog polja tj. MRF od *Markov random field*. Upotrebom predloženog modela u postupku rekonstrukcije definisaćemo nove regularizacione funkcije $\phi$. U nastavku ćemo dati predlog statističkog modela za reprezentaciju sliku u transformacionom domenu i predloge metoda rekonstrukcije.

## Predložene metode rekonstrukcije

Postoje dva pristupa modelovanju strukture retke reprezentacije: (1) u toku akvizicije primenom različito dizajniranih trajektorija odabiranja i (2) u fazi rekonstrukcije kroz unapređenu regularizaciju inverznog problema. U tezi ćemo se fokusirati na drugi pristup, i u nastavku će biti izložen predlog funkcije regularizacije koji inkorporira statistički model strukture koeficijenata retke reprezentacije signala.

Označimo sa $\mathbf{P}^{D\times N}$ transformaciju za retku reprzentaciju sig-

nala preko koeficijenata $\boldsymbol{\theta} = \mathbf{P}\mathbf{x} = \{\theta_1, ..., \theta_D\}$. Koeficijenat $\theta_i$ se smatra *značajnim* ako je njegova magnituda iznad nekog praga. Označimo sa skrivenom labelom $s_i \in \{0, 1\}$ koeficijenat $\theta_i$ da bismo razlikovali koeficijente po značajnosti: $s_i = 1$ ako je $\theta_i$ značajan i $s_i = 0$ ako nije. Određena konfiguracija $\mathbf{s} = \{s_1, ..., s_D\}$ se posmatra kao realizacija slučajnog Markovljevog polja $\mathbf{S} = \{S_1, ..., S_D\}$. Slika. 5 ilustruje celokupnu postavku problema i usvojen model. Svako merenje u $\mathbf{y}$ je dobijeno kao linearna kombinacija intenziteta svih $N$ piksela u slici $\mathbf{x}$ kroz operator $\mathbf{A}$. Retki koeficijenti $\boldsymbol{\theta}$ dobijaju se nakon primene operatora analize $\mathbf{P}$ nad slikom $\mathbf{x}$. Stoga svaki koeficijent $\theta_i$ je linearna kombinacija svih vrednosti piksela preko transformacije $\mathbf{P}$. Obrnuto, svaka vrednost piksela $x_i$ je dobijena kao linearna kombinacija svih koeficijenata $\theta_i$ korišćenjem operatora sinteze $\mathbf{P}^H$.

Označimo skup indeksa pozicija koji odgovara konfiguraciji $\mathbf{s}$ sa $\Omega_{\mathbf{s}} = \{i \in \mathcal{N} : s_i = 1\}$, uzimajući u obzira samo kad je $s_i = 1$, gde je skup indeksa svih pozicija označen sa $\mathcal{N} = \{1, 2, 3, ..., D\}$. Stoga, model za $\boldsymbol{\theta}$ koji odgovara konfiguraciji $\mathbf{s}$ definišemo kao:

$$\mathcal{M}_{\mathbf{s}} = \{\boldsymbol{\theta} \in \mathbb{C}^D : supp(\boldsymbol{\theta}) = \Omega_{\mathbf{s}}\}. \tag{28}$$

što dalje dovodi do sledećeg pristupa u estimaciji slike $\mathbf{x}$:

$$\min_{\mathbf{x} \in \mathbb{C}^N} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \quad \text{subject to} \quad \mathbf{P}\mathbf{x} \in \mathcal{M}_{\hat{\mathbf{s}}} \tag{29}$$

gde je $\hat{\mathbf{s}}$ estimacija najverovatnih pozicija značajnih koeficijenata retke reprezentacije $\boldsymbol{\theta} = \mathbf{P}\mathbf{x}$. Uslov $\mathbf{P}\mathbf{x} \in \mathcal{M}_{\hat{\mathbf{s}}}$ može ekvivalentno biti zamenjen sa $supp(\mathbf{P}\mathbf{x}) = \Omega_{\hat{\mathbf{s}}}$. Rešavanjem ovog problema obuhvata rešavanje sledećeg problema

$$\min_{\boldsymbol{\gamma} \in \mathbb{C}^D} \|\boldsymbol{\gamma} - \boldsymbol{\theta}\|_2^2 \quad \text{subject to} \quad \boldsymbol{\gamma} \in \mathcal{M}_{\hat{\mathbf{s}}} \tag{30}$$

za koji je rešenje $\hat{\boldsymbol{\gamma}}_H[\Omega_{\hat{\mathbf{s}}}] = \boldsymbol{\theta}[\Omega_{\hat{\mathbf{s}}}]$ i $\hat{\boldsymbol{\gamma}}_H[\bar{\Omega}_{\hat{\mathbf{s}}}] = 0$. Pošto $s_i \in \{0, 1\}$, rešenje može biti zapisano kao *Hadamard*-ov proizvod $\hat{\boldsymbol{\gamma}}_H[\Omega_{\hat{\mathbf{s}}}] = \boldsymbol{\theta} \circ \hat{\mathbf{s}}$.

Pretraga najverovatnije konfiguracije $\hat{\mathbf{s}}$ je izvršena korišćenjem kriterijuma maksimalne aposteriorne verovatnoće:

$$\hat{\mathbf{s}} = \underset{\mathbf{s}}{\operatorname{argmax}}\, P_{\mathbf{S}|\boldsymbol{\theta}}(\mathbf{s} \mid \boldsymbol{\theta}) = \underset{\mathbf{s}}{\operatorname{argmax}}\, p_{\boldsymbol{\theta}|\mathbf{S}}(\boldsymbol{\theta} \mid \mathbf{s}) P_{\mathbf{S}}(\mathbf{s}) \tag{31}$$

U praksi estimacija $\hat{\mathbf{s}}$ će se vršiti u svakoj iteraciji algoritma rekonstrukcije, počeći od trenutne, privremene estimacije vektora koeficijenata retke reprezentacije $\boldsymbol{\theta}$. Na osnovu estimiranog $\hat{\mathbf{s}}$ ćemo definisati

regularizacionu funkciju $\phi$ koju ćemo koristiti u postavci optimizacionog problema za rekonstrukciju datog u (26).

Združena verovatnoća $P_{\mathbf{S}}(\mathbf{s})$ Markovljevog slučajnog polja je Gibsova distribucija [Li 09], [Pižurica 02]

$$P_{\mathbf{S}}(\mathbf{s}) = \frac{1}{Z} e^{-H(\mathbf{s})/T} \tag{32}$$

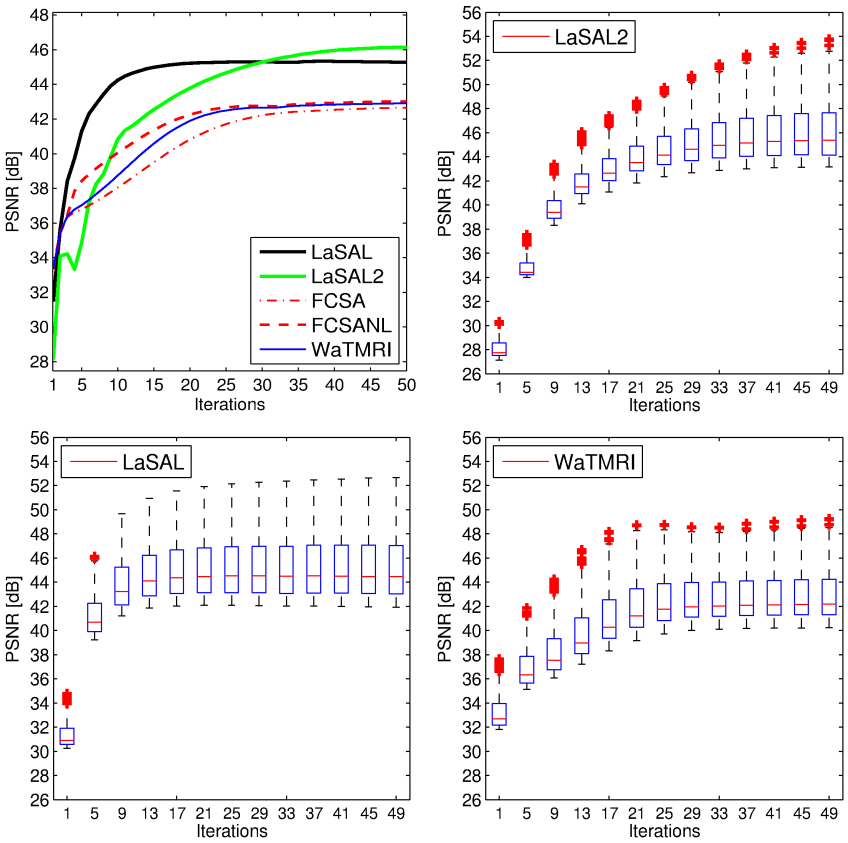
gde je funkcija energija $H(\mathbf{s})$ predstavljena preko sume potencijala klika uzimajući u obzir sve moguće klike: $H(\mathbf{s}) = \sum_{c \in \mathcal{C}} V_c(\mathbf{s})$. Na ovaj način je omogućena efikasna faktorizacija Gibsove distribucije koristeći potencijalne funkcije koje modeliraju lokalne interakcije. Za funkciju energije koristimo Izignov model kao u radu [Pižurica 11] koji je karakterisan parnim klikovima nad kojim su definisane potencijane funkcije. Izingov model, kao pojednostavljena verzija auto-logističkog modela, ima sledeću formu

$$H(\mathbf{s}) = \sum_i V_1(s_i) + \sum_{\langle i,j \rangle \in \mathcal{C}} V_2(s_i, s_j) \tag{33}$$

sa potencijalnim funkcijama definisanim na sledeći način

$$V_1(s) = \begin{cases} \alpha & s = 0 \\ -\alpha & s = 1 \end{cases}, \quad V_2(s,t) = \begin{cases} -\beta & s = t \\ \beta & s \neq t \end{cases} \tag{34}$$

Kondicionalni model $p_{\mathbf{\Theta}|\mathbf{S}}(\boldsymbol{\theta}|\mathbf{s})$ usvajamo iz radova [Pižurica 11, Pižurica 02]. Sa pretpostavkom o uslovnoj nezavisnosti imamo $p_{\mathbf{\Theta}|\mathbf{S}}(\boldsymbol{\theta}|\mathbf{s}) = \prod_i p_{\Theta_i|S_i}(\theta_i|s_i)$. Posmatrani koeficijenti su zašumljena verzija idealnih: $\theta = u+n$, gde $n$ označava komponentu šuma. Odabrali smo generalnu Laplasovu raspodelu za $p_U(u)$ i estimirali smo njene parametre iz histograma zašumljenih koeficijenata, znajući standardnu devijaciju šuma $\sigma$ [Pižurica 02, Simoncelli 97]. U praksi, $\sigma$ se pouzdano estimira iz praznog prostora blizu granica MR slike i skalira za svaki podband koeficijenata. Sa $T_h$ označavamo prag značajnosti za idealne koeficijente bez šuma ($u$ je značajan ako $|u| \geq T_h$). Povezujemo ovaj prag sa nivoom šuma na sledeći način, $T_h$ je samo deo $\sigma$ (u praksi 10%). Uslovne gustine verovatnoće $p_{U|S}(u|0)$ i $p_{U|S}(u|1)$ su dobijene skaliranjem centralnog dela ($|u| < T_h$) i repova ($|u| \geq T_h$) gustine raspodele $p_U(u)$ tako da integral obe funkcije gustine bude 1. Uslovne gustine verovatnoće zašumljenih koeficijenata $p_{\Theta|S}(\theta|s)$ su dobijeni iz $p_{U|S}(u|s)$. Za aditivni model šuma $\theta = u + n$ sa $n \sim N(0, \sigma)$, $p_{\Theta|S}(\theta|s)$ je konvolucija $p_{U|S}(u|s)$ sa $N(0, \sigma)$.




**Slika 6:** PSNR vrednosti dobijene iz 248 MR slika. Srednja vrednost PSNR (**Gore levo**) i distribucija PSNR vrednosti za LaSAL2 (**dole desno**), LaSAL (**dole levo**) i WaTMRI (**dole desno**). Rezultati su prezentovani na sledeći način: ivice svakog pravougaonika prezentuju $25^{th}$ i $75^{th}$ percentil dok centralna crvena linija predstavlja median. Sa zvezdicama su predstavljene ekstremne vrednosti.

Na osnovu predloženih modela definišemo sledeću regularizacionu funkciju koristeći model $\mathcal{M}_{\hat{\mathbf{s}}}$:

$$\phi(\boldsymbol{\theta}) = \begin{cases} 0, & \text{if } \boldsymbol{\theta} \in \mathcal{M}_{\hat{\mathbf{s}}} \\ \infty, & \text{if } \boldsymbol{\theta} \notin \mathcal{M}_{\hat{\mathbf{s}}} \end{cases} \tag{35}$$

koju ćemo koristiti u optimizacionoj postavci problema u (26). Rešavanjem optimizacionog problema sa predloženom funkcijom regularizacije ili sa kombinacijom predložene funkcije i totalne varijacije (TV) kao još jednog vida regularizacije, pomoću CSALSA metode (*constrained split*

**Algorithm 1** LaSAL

**Input:** $k = 0, \mu > 0, \mathbf{v}^{\{0\}}, \mathbf{w}^{\{0\}}, \mathbf{b}^{\{0\}}, \mathbf{c}^{\{0\}}$

**repeat**
  $\mathbf{r}^{\{k\}} = \mu(\mathbf{w}^{\{k\}} + \mathbf{c}^{\{k\}}) + \mathbf{A}^H(\mathbf{v}^{\{k\}} + \mathbf{b}^{\{k\}})$
  $\mathbf{x}^{\{k+1\}} = (\mu\mathbf{I} + \mathbf{A}^H\mathbf{A})^{-1}\mathbf{r}^{\{k\}}$
  $\mathbf{v}^{\{k+1\}} = \mathbf{\Psi}_{\iota_{E(\epsilon,\mathbf{I},\mathbf{y})}}(\mathbf{A}\mathbf{x}^{\{k+1\}} - \mathbf{b}^{\{k\}})$
  $\boldsymbol{\theta}' = \mathbf{P}(\mathbf{x}^{\{k+1\}} - \mathbf{c}^{\{k\}})$
  $\hat{\mathbf{s}} \leftarrow \mathit{MAP\text{-}support}\{\boldsymbol{\theta}'\}$
  $\mathbf{w}^{\{k+1\}} = \mathbf{P}^H(\boldsymbol{\theta}' \circ \hat{\mathbf{s}})$
  $\mathbf{b}^{\{k+1\}} = \mathbf{b}^{\{k\}} - (\mathbf{A}\mathbf{x}^{\{k+1\}} - \mathbf{v}^{\{k+1\}})$
  $\mathbf{c}^{\{k+1\}} = \mathbf{c}^{\{k\}} - (\mathbf{x}^{\{k+1\}} - \mathbf{w}^{\{k+1\}})$
  $k = k + 1$
**until** some stopping criterion is satisfied

**Algorithm 2** LaSAL2

**Input:** $k = 0, \mu_1, \mu_2 > 0, \mathbf{v}^{\{0\}}, \mathbf{w}^{\{0\}}, \mathbf{z}^{\{0\}}, \mathbf{b}^{\{0\}}, \mathbf{c}^{\{0\}}, \mathbf{d}^{\{0\}},$

**repeat**
  $\mathbf{r}^{\{k\}} = \mu_1(\mathbf{z}^{\{k\}} + \mathbf{c}^{\{k\}}) + \mathbf{A}^H(\mathbf{v}^{\{k\}} + \mathbf{b}^{\{k\}})$
  $\mathbf{x}^{\{k+1\}} = (\mu_1\mathbf{I} + \mathbf{A}^H\mathbf{A})^{-1}\mathbf{r}^{\{k\}}$
  $\mathbf{v}^{\{k+1\}} = \mathbf{\Psi}_{\iota_{E(\epsilon,\mathbf{I},\mathbf{y})}}(\mathbf{A}\mathbf{x}^{\{k+1\}} - \mathbf{b}^{\{k\}})$
  $\mathbf{z}' = \frac{1}{(\mu_1+\mu_2)}\big(\mu_1(\mathbf{x}^{\{k+1\}} - \mathbf{c}^{\{k\}}) + \mu_2(\mathbf{w}^{\{k\}} + \mathbf{d}^{\{k\}})\big)$
  $\mathbf{z}^{\{k+1\}} = \mathbf{\Psi}_{\mathrm{TV}}(\mathbf{z}'; \mu_1 + \mu_2)$
  $\boldsymbol{\theta}' = \mathbf{P}(\mathbf{z}^{\{k+1\}} - \mathbf{d}^{\{k\}})$
  $\hat{\mathbf{s}} \leftarrow \mathit{MAP\text{-}support}\{\boldsymbol{\theta}'\}$
  $\mathbf{w}^{\{k+1\}} = \mathbf{P}^H(\boldsymbol{\theta}' \circ \hat{\mathbf{s}})$
  $\mathbf{b}^{\{k+1\}} = \mathbf{b}^{\{k\}} - (\mathbf{A}\mathbf{x}^{\{k+1\}} - \mathbf{v}^{\{k+1\}})$
  $\mathbf{d}^{\{k+1\}} = \mathbf{d}^{\{k\}} - (\mathbf{z}^{\{k+1\}} - \mathbf{w}^{\{k+1\}})$
  $\mathbf{c}^{\{k+1\}} = \mathbf{c}^{\{k\}} - (\mathbf{x}^{\{k+1\}} - \mathbf{z}^{\{k+1\}})$
  $k = k + 1$
**until** some stopping criterion is satisfied

*augmented Lagrangian metode* izvodimo dva algoritma za rekonstrukciju slike: LaSAL i LaSAL2 predstavljeni u algoritmima 1 i 2.

Oba algoritma sadrže korak vezan za estimaciju pozicija značajnih koeficijenata $\hat{\mathbf{s}} \leftarrow \mathit{MAP\text{-}support}\{\boldsymbol{\theta}'\}$ na osnovu kojeg se posle vrši postavljanje svih koeficijenata koji nisu na pozicijama gde je $s_i = 1$ na vrednost nula.

Na slici 6 vidimo rezultat eksperimenta izražen u PSNR (Peak

**Algorithm 3** FCLaTV

**Input:** $k = 1, \mu, \tau_1, \tau_2, t^{\{1\}} = 1, \mathbf{y}, \mathbf{r}^{\{1\}} = \mathbf{x}^{\{0\}}$

**repeat**
  $\mathbf{x}_g = \mathbf{r}^{\{k\}} - \mu \mathbf{A}^H(\mathbf{A}\mathbf{r}^{\{k\}} - \mathbf{y})$
  $[\phi_{\hat{\mathbf{s}}}, \hat{\mathbf{s}}] = \textit{MAP-support}\{\mathbf{P}\mathbf{x}_g\}$
  $\mathbf{x}_1 = \mathbf{P}^H(\mathbf{prox}_{\mu}(\tau_1 \phi_{\hat{\mathbf{s}}})(\mathbf{P}\mathbf{x}_g))$
  $\mathbf{x}_2 = \mathbf{G}^H(\mathbf{prox}_{\mu}(\tau_2 \psi_{\mathrm{TV}})(\mathbf{G}\mathbf{x}_g))$
  $\mathbf{x}^{\{k\}} = (\mathbf{x}_1 + \mathbf{x}_2)/2$
  $t^{\{k+1\}} = \left(1 + \sqrt{1 + 4(t^{\{k\}})^2}\right)/2$
  $\mathbf{r}^{\{k+1\}} = \mathbf{x}^{\{k\}} + \frac{t^{\{k\}}-1}{t^{\{k+1\}}}(\mathbf{x}^{\{k\}} - \mathbf{x}^{\{k-1\}})$
  $k = k + 1$
**until** some stopping criterion is satisfied
**return** $\mathbf{x} = \mathbf{x}^{\{k\}}$

**Algorithm 4** The proposed algorithm: GreeLa

**Input:** $k = 1, \mathbf{y}, \mathbf{x}^{\{0\}}, \mathbf{t} = \mathbf{0}$

**repeat**
  $\mathbf{r}^{\{k\}} = \mathbf{y} - \mathbf{A}\mathbf{x}^{\{k-1\}}$
  $\mathbf{x}_t^{\{k\}} = \mathbf{A}^H \mathbf{r}^{\{k\}} + \mathbf{x}^{\{k-1\}}$
  $\boldsymbol{\theta}_t^{\{k\}} = \mathbf{P}\mathbf{x}_t^{\{k\}}$
  $\mathbf{s}^{\{k\}} = \textit{MAP-support}\{\boldsymbol{\theta}_t^{\{k\}}\}$
  $\mathbf{t} = \mathbf{0}; \quad \mathbf{t}[\mathbf{s}^{\{k\}} = 1] = \boldsymbol{\theta}_t^{\{k\}}[\mathbf{s}^{\{k\}} = 1]$
  $\boldsymbol{\theta}^{\{k\}} = \mathbf{t}, \mathbf{x}^{\{k\}} = \mathbf{P}^H \boldsymbol{\theta}^{\{k\}}$
  $k = k + 1$
**until** Maximum iterations or $\|\mathbf{r}^{\{\mathbf{k}\}}\| \leq \textit{threshold}$

Signal to Noise Ratio) meri gde je izvršena rekonstrukcija 248 MR slika iz 50% merenja. LaSAL2 postiže značajno veći PSNR nego referentne metode: unapređene je više od 3.5 dB. LaSAL metoda koja ne uključuje TV regularizaciju, postiže bolje rezulate od FCSANL i WaTMRI za oko 2.4 dB, i dostiže svoj najveći PSNR u znatno manje iteracija nego referentne merode. Na istoj slici smo prikazali i rezultujuće distribucije PSNR vrednosti po iteraciji. Može se videti da nakon 5 iteracija LaSAL postiže znatno unapređene u vrednostima PSNR u odnosu na sve referentne metode, dok LaSAL2 postiže bolje rezultate nego LaSAL posle 30 iteracija.

Predložen pristup regularizacije sa estimacijom pozicija značajnih koeficijenata se može koristiti i u drugom iterativnom algoritmu

za rekonstrukciju. Pored CSALSA metode koristili smo i FCSA (*fast composite splitting algorithm*) metodu gde smo zajedno sa zajedno sa TV regularizacionom funkcijom koristili modifikaciju predloženog pristupa regularizacije na osnovu Markovljevog modela. Koristili smo anizotropni Izingov model sa automatskom estimacijom parametara za razliku od izotropnog Izingovog modela korišćenog u razvoju LaSAL i LaSAL2 metode. Razvijen metod, nazvan FCLaTV koristi sledeće pravilo odlučivanja i regularizacije koeficijenata:

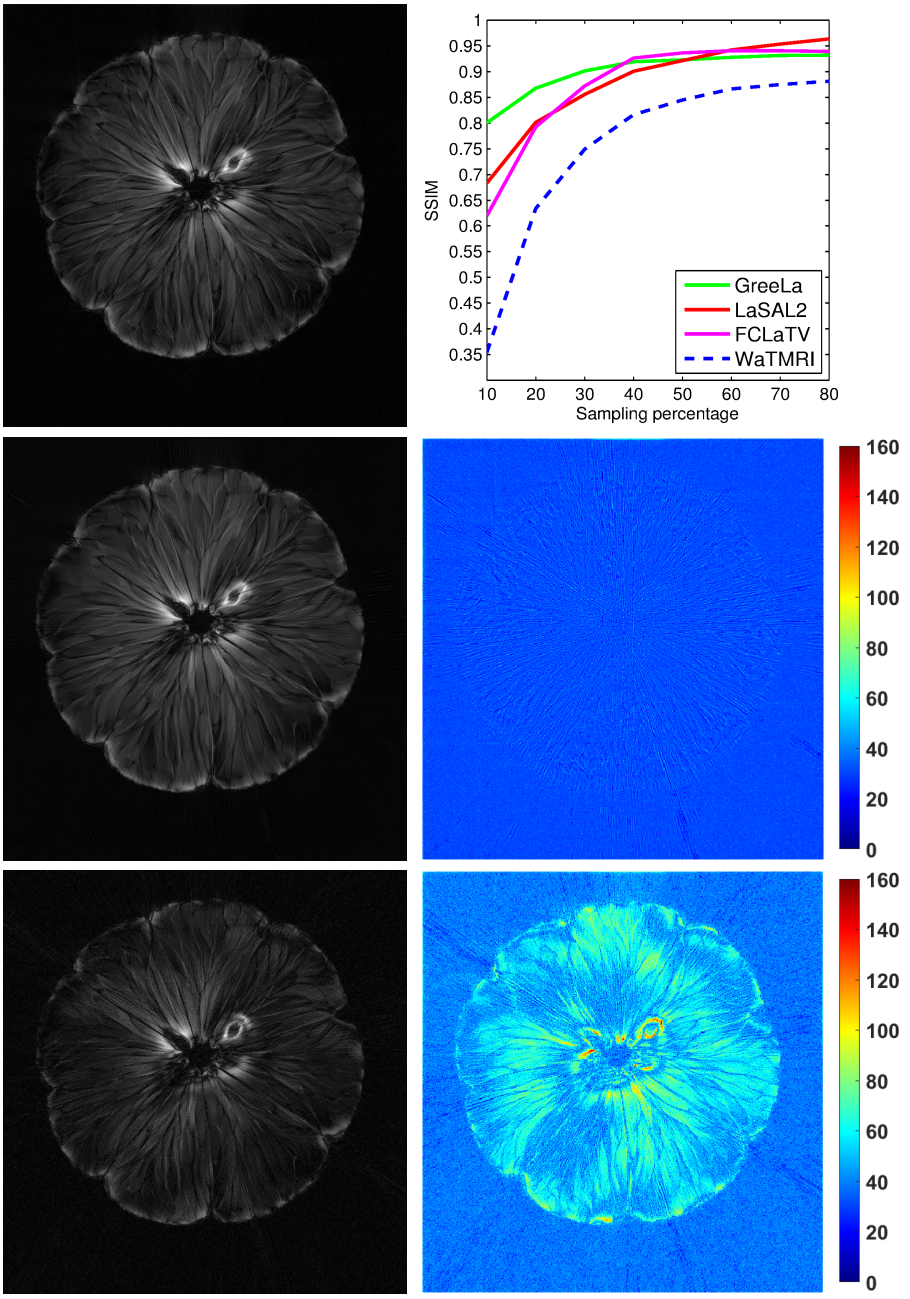
$$\hat{\theta}_{\hat{s}=1} = \begin{cases} B \cdot \operatorname{sgn}(\theta_g), & |\theta_g| \leq \frac{\mu}{b} + B \\ \theta_g - \frac{\mu}{b} \cdot \operatorname{sgn}(\theta_g), & |\theta_g| > \frac{\mu}{b} + B \end{cases}$$
$$\hat{\theta}_{\hat{s}=-1} = \begin{cases} 0, & |\theta_g| \leq \frac{\mu}{b} \\ \theta_g - \frac{\mu}{b} \cdot \operatorname{sgn}(\theta_g), & \frac{\mu}{b} < |\theta_g| \leq \frac{\mu}{b} + B \\ B \cdot \operatorname{sgn}(\theta_g), & |\theta_g| > \frac{\mu}{b} + B \end{cases} \tag{36}$$

gde $\theta_g$ označa privremeno estimirane koeficijente, $\mu$ regularizacioni parametetar a $b, B$ parametre koji se estimiraju iz histograma zašumljenih koeficijenata. Prikaz algoritma dat je u algoritmu 3.

Pored iterativnih metoda izvedenih iz optimizacione definicije problema rekonstrukcije, testirali smo i pohlepni algoritam rekonstrukcije sa predloženom metodom regularizacije. Oslonili smo se na LaMP (*lattice matching pursuit*) algoritam i proširili smo ga slučaj rekonstrukcije slika koje imaju retku reprezentaciju u transformacionom domenu. Izvedeni algoritam GreeLa (*greedy lattice*) dat je u algoritmu 4.

Urađen je eksperiment na realnim podacima dobijenim skeniranjem voća pomelo. Na slici 7 su data poređena performansi izražena u SSIM (structure similarity measure) meri za sledeće algoritme: FCLaTV, LaSAL2, GreeLa i WaTMRI. Date su rekonstruisane slike korišćenjem GreeLa i WaTMRI metoda i njihove greške radi vizualnog poređenja. Za sve procente odabiranja, predložene metode postižu bolje performanse nego WaTMRI. GreeLa metod postiže viši SSIM nego LaSAL2 i FCLaTV za procenat odabiranja ispod 40%. Za više procente odabiranja LaSAL2 i FCLaTV sa svojim kompleksnijim iterativnim procedurama i regularizacijom koja uključuje i TV postižu bolje rezultate nego GreeLa.

Od predloženih metoda izdvojili smo LaSAL2 i proširili ga na slučaj rekonstrukcije MR slike iz merenja dobijenih sa više kalemova tokom skeniranja MR uređajem. Izvršene promene na LaSAL2 metodi obuhvataju korišćenje merenja sa svih kalemova istovremeno i regularizaciju kompozitne (finalne) slike a ne slika od svakog kalema posebno.




**Slika 7:** Pomelo eksperiment.. **Prva kolona odgore na dole**: referentna slika dobijena iz 100% merenja, rekonstrukcija iz 20 % merenja koristeći GreeLa i WaTMRI metode redom. **Druga kolona od gore na dole**: Dobijena vrednost SSIM za različite procente odabiranja, praćena greškama rekonstrukcije koje odgovaraju metodama iz prve kolone.

| Metode | SSIM | NRMSE |
|---|---|---|
| P-LORAKS | 0.88 / 0.92 | 0.07 / 0.06 |
| SENSE-LORAKS | 0.83 / 0.87 | 0.08 / 0.07 |
| Predložena metoda | **0.95** / **0.96** | **0.04** / **0.03** |

**Tabela 1:** Rekonstrukcija T1 MR slike mozga iz rada [Kim 17] korišćenjem nepotpunih merenja iz 4 kalema sa uniformnom / slučajnom trajektorijom. Poređenje sa P-LORAKS i SENSE-LORAKS metodama sa predloženim metodom je dato u SSIM i NRMSE (*normalized root-mean-squared-error*) meri.

U tabeli su data poređenja sa trenutno najnaprednijim algoritmima za rekonstrukciju slike iz merenja sa više kalemova.

## Zaključak

Predložena regularizacija koja koristi Markovljevo slučajno polja za modelovanje retke reprezentacije signala pokazala se izuzetno uspešno u problemima rekonstrukcije MR slike iz nepotupnih merenja. U različitim iterativnim algoritmima testirali smo predloženi model kroz postupak regularizacije tokom rekonstrukcije slike. U kombinaciji sa TV regularizacijom dobijeni su najbolje performanse u rekonstrukciji na svim testiranim iterativnim procedurama. Na kraju algoritam je proširen i na problem rekonstrukcije iz merenja dobijenih sa više kalemova. Svi rezultati pokazuju superiornost predloženog metoda regularizacije u odnosu na trenutno state-of-the-art metode u literaturi.

# Summary

Medical diagnosis increasingly relies on various technical devices, including imaging sensors, which incorporate comprehensive knowledge of various fields of physics, mathematics, and electrical engineering. These devices often provide essential information to medical experts about anatomical or functional features of interest for diagnosis or follow-up during the treatment. In modern medical imaging, magnetic resonance imaging (MRI) plays a fundamental role due to high spatial resolution and non-ionizing nature. MRI is based on the principles of nuclear magnetic resonance (NMR), a spectroscopic technique used by scientists to obtain microscopic chemical and physical information about molecules. The origin of the NMR signal lays in electromagnetic radiation emitted from the hydrogen nuclei in the human body when they are stimulated by a radio-frequency (RF) pulse while keeping them in a strong static magnetic field. Since a human body is approximately 63% composed of hydrogen atoms, MRI primarily images the NMR signal from the hydrogen nuclei. Sometimes a chemical contrast agent is introduced into the body to improve the visibility of internal body structures by changing the NMR signal strength. Spatial variations in the phase and frequency of the radiofrequency energy being absorbed and emitted by the imaged object are what we call MR images. In this thesis, we will focus our research on signal and image processing aspects of MRI to further improve the achievable image quality and/or to reduce the image acquisition time.

MRI scans vary from 20 to 90 minutes, depending on the size of the area being scanned and whether or not a contrast agent is added. A typical MRI scan of the abdomen often takes an hour. Some types of examinations require patients to repeatedly hold their breath during the scanning session. The MRI scanner produces loud repetitive knocking sounds at certain times during the scanning procedure. This is a consequence of the electric current in the scanner coils being turned on and off. Due to the volume of these sounds, it is recommended that patients wear earplugs or headphones which are provided during the scanning session. Some imaging sites provide an airplane-like audio system for those being

imaged. These systems provide some noise suppression, and also mask out the unpleasant sound with music. Not very often some individuals experience magnetophosphenes, i.e., visual sensations of flashes of light on the retina. Other sensations that have been reported by some individuals during their magnetic resonance scan include vertigo or dizziness, a metallic taste, and nausea. Patients who are anxious or nervous about enclosed spaces can be given medication prior to the MRI to help to make the procedure slightly more bearable. The reduction of scanning time and therefore exposure to unpleasant and uncomfortable circumstances during MRI scanning is a primary goal that must go along with achieving good image quality. Another beneficial advantage from faster scanning is a higher throughput of examination and therefore reduction of waiting queues. Since the acquisition time of MR image is directly proportional to the number of collected measurements, and since better image quality is provided with more measurements, these two requirements are contradictory and compromise is needed.

Two primary approaches to accelerate the acquisition are: i) hardware design which allows simultaneous acquisition from multiple coils and ii) the use of nonlinear image reconstruction procedures that recover images from partially sampled data. In this thesis, we developed improved nonlinear reconstruction algorithms, that take as an input fewer samples than in traditional MRI acquisitions to decrease the image acquisition time. The underlying theory that allows for image reconstruction from undersampled data is commonly referred to as compressed sensing (CS). In CS theory, general conditions have been established regarding the properties of signals and measurements design under which nearly perfect reconstruction of signals is possible from measurements sampled below the Nyquist rate. General MR images, although not ideally sparse, belong to the so-called compressible signals, for which the CS theory provides provable reconstruction guarantees and works well in practice.

The recovery of MR images from undersampled measurements is an ill-posed problem and thus some sort of regularization must be employed which will help in finding the right solution among all possible solutions determined by a problem formulation. Commonly employed types of regularization for inverse imaging problems, including MRI reconstruction, are total variation (TV) and sparsity-enforcing constraints in wavelet-like domains. While promoting piece-wise smooth solutions, these regularizations do not capture the global structure of image edges and textures at different scales. This leaves much space for improving

the existing image reconstruction algorithms by introducing new types of regularization. One of the main contributions of this thesis is a new statistical model for sparse signal representation from which a new regularization rule is derived. The proposed model enforces the structured sparsity signal representation which corresponds to existing structures in images like edges and textures. A Markov Random Field (MRF) model, in particular, Ising model, is involved as a prior for the signal support taking into account correlations between the sparse image coefficients and therefore restricting the space of possible solutions for the reconstructed image. This restriction (constraint) is included in the reconstruction procedure as a regularization function that guides the pursuit of the solution.

We developed two types of methods for MRI reconstruction with the proposed MRF based regularization. The first type of methods belongs to the class of greedy based algorithms. A greedy algorithm builds a specific candidate solution incrementally. The aspect of a greedy algorithm that makes it 'greedy' is how it chooses from among the different ways of incrementing the current partial solution. In general, different choices are ordered according to some criteria, and the best choice according to these criteria is taken. Thus the algorithm builds the solution by always taking the step that appears to be most promising at that moment. Though there are many problems for which greedy strategies do not produce optimal solutions, when they do, they tend to be quite efficient. We proposed a greedy method that utilizes the lattice structure of an MRF model in obtaining a new image estimate through iterations. We named this method GreeLa from greedy lattice regularization and tested it on real Cartesian and non-Cartesian generated MR data. Inclusion of MRF based regularization through a greedy reconstruction framework dramatically increased the speed of the search for a solution and improved reconstruction performances compared to the other state-of-the-art methods.

The second type of methods that we developed belong to the class of optimization-based algorithms for inverse imaging problems. The optimization-based approach proposes an image estimate as a solution of an optimization problem where the objective function might be of convex or non-convex nature depending on the type of regularization terms that are involved. Therefore the image estimate is treated as a local or, in the best case, the global optimum of the objective function. The search for the optimum of the objective function, which might be under some constraints, is conducted with an iterative algorithm whose

steps are derived from the mathematical formulation of the objective function and constraints themselves. Since we formulate the MRI reconstruction problem as a constrained optimization problem, we solved it using a constrained split augmented Lagrangian shrinkage algorithm (C-SALSA) which is demonstrated to be a particularly successful approach for various inverse imaging problems such as denoising, deblurring, and reconstruction. C-SALSA solves a constrained optimization problem by transforming it into an unconstrained one which is further transformed into a different constrained problem using the variable splitting operation. Then the obtained constrained problem is solved by the augmented Lagrangian approach. From this, a regularization rule is obtained in a form of Moreau proximal mapping associated with regularization functions (terms) used in the objective function. We improve the C-SALSA method with the proposed regularization function based on an MRF signal prior to infer the support (locations) of the important information-bearing coefficients. Besides MRF based regularization, we also consider a compound regularization of total variation (TV) alongside the proposed MRF. As a result, we derived two particular algorithms following the same principles as in the derivation of the C-SALSA framework. Since both of the algorithms involve a split augmented Lagrangian approach and a lattice-based regularization from MRF prior, we refer to the algorithm as LaSAL (when using MRF regularization alone) and LaSAL2 in the case of compound MRF-TV regularization. Both of the proposed methods are tested on datasets obtained by undersampling measurements using various sampling trajectories. They demonstrated improved reconstruction performances compared to traditional and related state-of-the-art methods in the field.

Further on, we analyzed the proposed regularization functions in LaSAL and LaSAL2 in a different optimization framework. We use a fast composite splitting algorithm (FCSA) as an iterative reconstruction procedure where we incorporate a compound (MRF prior + TV) regularization. In FCSA, a Moreau proximal map for the compound regularization is approximated by averaging results obtained using the proximal maps for each regularization (MRF and TV) separately. Another novelty that we introduce is the usage of an anisotropic version of the Ising model as the signal prior instead of the isotropic version utilized in LaSAL and LaSAL2. This improvement, which brings additional parameters in the MRF model, is accompanied by the proposed efficient estimation of the MRF parameters through algorithm iterations. Finally, a new soft-thresholding regularization based rule is derived using the

anisotropic MRF model and estimated signal coefficient's support mask. The proposed method, named FCLaTV (fast composite lattice and TV), achieved similar performances as our other method with compound regularization LaSAL2. The automatic procedure for the estimation of MRF parameters incorporated in FCLaTV is of crucial importance for practical usage of algorithm. While LaSAL2 has better utilization of more measurements compared to FCLaTV which is reflected in slightly better reconstruction, FCLaTV on the other side converges faster to a stable solution without initial MRF parameters setup.

Finally, to get closer to the real situation in practice, we consider MR image reconstruction from undersampled multi-coil measurements. Most of the reported methods for the reconstruction from multi-coil data, or so-called parallel MRI (pMRI), address the problem where the measurements from different coils are fully sampled. Instead, we consider a compressed sensing approach to parallel MRI, i.e., CS-pMRI, where the measurements coming from each coil are undersampled using the same sampling trajectory. We create a joint framework for the reconstruction of a composite MR image from all undersampled coil measurements using the estimated coil-sensitivity profiles. The developed algorithm is built on the LaSAL2 framework with extensions made to accommodate the inputs coming from multiple coils. The experimental results show improvements over the best performing CS-pMRI methods from the literature.

In terms of the published results, the work presented in this thesis resulted in a journal paper in the IEEE Transactions on Medical Imaging and four conference papers for all of which the author of this thesis is the first author. Another journal paper is currently being finalized for submission.

# Contents

# Symbols and acronyms used in this thesis

## Symbols

| | |
|---|---|
| $I$ | Proton angular moment (spin) |
| $\mu$ | Proton magnetic moment |
| $\omega_L$ | Angular Larmor frequency |
| $\mathbf{M}$ | Macroscopic magnetization vector |
| $\mathbf{B}_0$ | External static magnetic field |
| $\mathbf{B}_1$ | Dynamic RF magnetic field |
| $T_1$ | Longitudinal relaxation time |
| $T_2$ | Transverse relaxation time |
| $\mathbf{G}$ | Linear magnetic gradient field |
| $s(t)$ | MRI signal |
| $\boldsymbol{\kappa}(t)$ | $\mathscr{k}$-space trajectory |
| $\boldsymbol{\rho}(x, y)$ | continuous MR image |
| $\mathbf{A}$ | Fourier encoding matrix |
| $\mathbf{y}$ | $\mathscr{k}$-space measurements |
| $\mathbf{x}$ | image under reconstruction |
| $\mathbf{P}$ | sparsifying transform |

## Acronyms

| | |
|---|---|
| NMR | Nuclear Magnetic Resonance |
| MRI | Magnetic Resonance Imaging |
| RF | Radio Frequency |
| SE | Spin Echo |
| FID | Free Induction Decay |
| TR | Repetition Time |
| FOV | Field Of View |
| PGM | Probabilistic Graphical Model |

| | |
|---|---|
| MRF | Markov Random Field |
| GRF | Gibbs Random Field |
| HMT | Hidden Markov Tree |
| CS | Compressed Sensing |
| RIP | Restricted Isometry Property |
| K-RIP | K-Restricted Isometry Property |
| CS-MRI | Compressed Sensing - Magnetic Resonance Imaging |
| BP | Basis Pursuit |
| BPDN | Basis Pursuit with Denoising |
| LASSO | Least Absolute Shrinkage and Selection Operator |
| IHT | Iterative Hard Thresholding |
| IST | Iterative Soft Thresholding |
| CoSaMP | Compressive Sampling Matching Pursuit |
| SP | Subspace Pursuit |
| LaMP | Lattice Matching Pursuit |
| TwIST | Two step Iterative Soft Thresholding Algorithm |
| FISTA | Fast Iterative Shrinkage-Thresholding Algorithm |
| SpaRSA | Sparse Reconstruction by Separable Approximation |
| FFT | Fast Fourier Transform |
| DFT | Discrete Fourier Transform |
| IDFT | Inverse Discrete Fourier Transform |
| DWT | Discrete Wavelet Transform |
| NDWT | Non-decimated Discrete Wavelet Transform |
| SIDWT | Shift Invariant Discrete Wavelet |
| PBDW | Patch Based Directional Wavelet |
| TV | Total Variation |
| PANO | Patch-Based Nonlocal Operators |
| C-SALSA | Constrained Split Augmented Lagrangian Shrinkage Algorithm |
| MAP | Maximum a Posteriori Probability |
| ML | Maximum Likelihood |
| MMSE | Minimum Mean Squared Error |
| ICM | Iterative Conditional Modes |
| LBP | Loopy Belief Propagation |
| MCMC | Markov Chain Monte Carlo |
| GreeLa | Greedy Lattice regularization |
| LaSAL | Lattice Split Augmented Lagrangian |
| ADMM | Alternating Directional Method of Multipliers |
| PSNR | Peak Signal to Noise Ratio |
| SSIM | Structure Similarity Measure |
| RLNE | Relative $\ell_2$ Norm Error |

| | |
|---|---|
| LASAL | Lattice Split Augmented Lagrangian |
| FCLATV | Fast Composite Lattice and TV |
| LORAKS | LOw RAnk modelling of local K-Space neighborhoods |
| SPIRIT | Iterative Self-consistent Parallel Imaging Reconstruction |
| ESPIRIT | Eigenvector-based implementation of SPIRiT |
| GRAPPA | Generalized Autocalibrating Partially Parallel Acquisitions |
| SENSE | SENSitivity Encoding |
| CG | Conjugate Gradient |

# 1

# Introduction

*Which fight is worth such sacrifices?*
*The one with which our mind and heart agree.*
—Meša Selimović

Magnetic resonance imaging (MRI) became one of the crucial diagnostic tools in modern medicine as it produces high-resolution images without ionizing radiation [Kuperman 00, Liang 00]. MRI scanners use a strong static magnetic field and radio-frequency (RF) wave emission to generate the image of the human anatomy [Rabi 38, Bloch 46, Purcell 46, Lauterbur 73]. Compared to the other medical imaging systems, MRI has several advantages. Among them, the principal advantage is superior contrast resolution, i.e., the ability to image differences among low-contrast tissues. Contrast resolution allows visualization of soft tissue with similar characteristics, such as liver-spleen or white matter-gray matter. When defining protocols of MRI there are many parameters whose proper tuning allows obtaining comprehensive imaging information for the area under observation. This flexibility in parameter choices is very important to optimize contrast resolution for visualization of various anatomical and disease states and to optimize spatial resolution concerning the scanning time. This is the case when imaging both stationary as well as non-stationary tissue such as flowing blood. An additional advantage to MRI is the ability to obtain direct transverse, sagittal, coronal, and oblique plane images. With this ability, a large data set is acquired during a single imaging sequence. This information helps in reconstructing any anatomical plane with excellent sensitivity which means that subtle differences in anatomy can be detected.

However, image acquisition is time-consuming and may last more than one hour depending on the type of examination which has been undertaken. Since MRI scanners produce loud and unpleasant sounds

during imaging, when the magnetic field gradients are turned on and off, it is recommended that patients wear earplugs or headphones. Some types of examinations require patients to repeatedly hold their breath during the scanning session. Some patients experience visual sensations of flashes of light on the retina during the examination. Those sensations called magnetophosphenes were first reported by the French physiologist Jaques-Arséne d'Arsonval in 1896. They are caused by induced electric currents in the retina in situations when the patient is moving through a static magnetic field, or when is stationary in the changing magnetic field. A narrow enclosed space in the MRI device, where the patient is positioned during scanning, can cause anxiety. In certain situations, medications are administered before the MRI scan to make the procedure slightly more bearable to the patient. The reduction of scanning time and thereby reducing exposure to unpleasant and uncomfortable circumstances during MRI scanning has been a primary goal of many recent studies and this PhD research. Reaching this goal must go along with keeping or improving the currently achievable MRI image quality and maintaining all the advantages that MRI brings compared to the other medical imaging systems.

To better understand how MRI scanning can be sped up, we start from the explanation of what kind of measurements MRI device gathers during scanning. We introduce the notion of spatial frequency domain or so-called $k$-space, which is crucial for MRI because all but the very earliest methods of MRI (such as the sensitive point method) scan the spatial frequency information about the image, and not the image plane itself. This means that the measurements obtained with an MRI device are the values of the components of a different representation of the image, so-called spatial frequency representation, in contrast to the standard spatial representation given by digital cameras. Hence the final image is obtained after the application of iterative reconstruction procedure which involves inverse Fourier transform to relate spatial frequency components with spatial location information, i.e., pixel values in the image. The number of acquired measurements with predefined sampling trajectories (schemes) in the $k$-space influences the overall image quality. Physical constraints of the MRI hardware are inherently the cause of the slow data acquisition process and limiting factor in the design of sampling trajectories. Therefore, there are two approaches to accelerate the acquisition process: i) hardware design which allows simultaneous acquisition from multiple coils, and ii) the use of nonlinear image reconstruction procedures that recover images from undersampled measure-

ments. The second approach leads to an ill-posed image reconstruction problem since measurements are sampled below the Shannon/Nyquist rate [Kotelnikov 33, Nyquist 28, Shannon 49, Whittaker 15]. In this thesis, we consider the second approach alone, where the measurements obtained from the single-coil are undersampled using differently designed sampling trajectories in $\mathcal{k}$-space, and in combination with the first approach when multiple coils are used.

Recently emerged compressed sensing or compressive sampling theory (CS) defines the necessary conditions under which the exact reconstruction of signals from undersampled measurements is possible making use of nonlinear algorithms. These conditions are related to the nature of the signal under reconstruction expressed through signal properties such as compressibility or sparseness, as well as through the properties of the sensing (acquisition) procedure. Signal compressibility or sparseness is not usually expressed in its canonical representation but among its coefficients in sparsifying, e.g., wavelet-like domain. Nonlinear reconstruction algorithms typically employ regularization which refers to a process of introducing certain constraints in order to solve an ill-posed problem. The rationale behind the regularization can be traced back to Occam's razor statement that simpler solutions are preferred to more complex ones, other things being the same. From a Bayesian point of view in MR reconstruction problems, regularization corresponds to imposing certain prior distributions on the signal model which limits the scope of possible solutions. Therefore, the involved regularization enforces the desired signal properties of a solution obtained by a nonlinear algorithm.

In CS-MRI reconstruction problems, the undersampled measurements acquired using one or more coils in the MRI device are taken as an input to the nonlinear reconstruction algorithm. Regularization, as a priori information about the desired solution, alleviates the ill-posedness of the problem. Mathematically this is formulated by adding a term in the objective function with whose solving we obtain the reconstructed image. Various nonlinear algorithms have been proposed so far in literature for this kind of problem [Fessler 10, Ying 04, Goldstein 09]. Commonly employed regularization are $\ell_1$ norm or total variation (TV) or a combination of these two imposed on the signal itself or its sparse representation [Lustig 07, Lustig 08, Ma 08, Yang 10, Qu 10, Huang 11b, Aelterman 11, Smith 13, Pejoski 15, Liu 16]. Regularization types are often selected in such a manner to induce some expected properties of the solution. These properties are something that we a priori know or assume about the image under reconstruction, such as sparseness

of image coefficients or piece-wise smoothness of the image itself. Recently, encoding the structure, i.e., the spatial clustering of the important MRI coefficients has been employed as a form o regularization both during the acquisition process [Adcock 14] and in the reconstruction [Pižurica 11, Chen 14, Chen 12].

In this thesis, the main focus will be on developing signal processing algorithms based on CS theory for the reconstruction of MR images from partial data, namely CS-MRI reconstruction. In [Pižurica 11] authors demonstrate a huge potential of lattice split-Bregman (LaSB) method in CS-MRI reconstruction. The main novelty that LaSB introduced is encoding the spatial structure with Markov random field (MRF) statistical model. Motivated by the presented results in [Pižurica 11] we propose a more general MRF model for the representation of spatial context among image coefficients, which reliably corresponds to the statistics of MR images. Based on the new MRF-based signal model, we derive new types of regularization for CS-MRI reconstruction. We thoroughly analyze suitable nonlinear procedures for image recovery where we incorporate derived MRF-based regularization. We develop two types of reconstruction algorithms: greedy and optimization-based. We evaluate the performances of the developed algorithms in terms of reconstruction quality and computational complexity under different sampling trajectories. A Cartesian and non-Cartesian grid are considered for a domain of sampling trajectories. The non-Cartesian grid additionally aggravates the reconstruction problem since it requires interpolation of measurements on the Cartesian grid before the application of the inverse Fourier transform. Obtained results are compared to the results of the state-of-the-art methods in the field of MRI reconstruction from single or multi-coil measurements. Improvements that are obtained in the results greatly justify introducing the MRF based regularization in nonlinear algorithms for MRI reconstruction and strongly motivate further research in this direction for other medical image modalities as well.

## 1.1 Main contributions

The main contributions of this thesis are:

- We develop an efficient method for MRI reconstruction from partial Fourier data making use of a Markov Random Field (MRF) prior for the support configurations of sparse coefficients. To our knowledge, this is the first elaborate study on CS-MRI with MRF

priors, although the potentials of such an approach were earlier demonstrated with a heuristic method named lattice split-Bregman (LaSB) in [Pižurica 11]. Compared to LaSB, the new method employs a different optimization technique, a more general MRF prior, and achieves consistently better results.

- We extend a constrained split augmented Lagrangian shrinkage algorithm (C-SALSA) of [Afonso 11] with an MRF prior. In particular, we introduce a new regularization step which admits support configurations favored by the prior model. The resulting algorithm, coined lattice split augmented Lagrangian (LaSAL) outperforms consistently C-SALSA.

- We develop a variant of the proposed method with compound regularization (MRF prior + TV norm) which further improves the reconstruction performance. A thorough evaluation is performed on MRI data sets acquired on Cartesian and non-Cartesian grids for which different undersampling strategies are simulated. For the radially acquired $k$-space data on the non-Cartesian grid we perform undersampling based on golden ratio profile spacing in order to reduce the inevitable error in interpolation step on the Cartesian grid. MRF-based CS-MRI methods demonstrate a clear improvement compared to alternative methods.

- We extend our MRF signal model to an anisotropic version and propose an efficient estimation of its parameters. Based on the more general MRF prior, we propose a soft-thresholding version of the regularization function. With estimated signal support in the sparse domain used to separate image coefficients into two classes, significant and insignificant, different soft-thresholding rules are derived for each class. The proposed MRF regularization, together with TV-norm, is integrated and tested under the fast composite splitting algorithm (FCSA) framework which showed good performances in MRI image recovery problems. Through experiments, we demonstrate the improvements obtained in MRI reconstruction using the FCSA method with proposed regularization, as compared to its original version. We reduce the number of parameters that need to be specified before the algorithm starts by involving the MRF parameter estimation procedure through iterations.

- We developed a method for MRI reconstruction from undersampled multi-coil measurements, in the so-called CS-pMRI setting.

Instead of reconstructing each coil image separately and then combining them to get the final MR composite image like most of the state-of-the-art algorithms in parallel MRI (pMRI) setting did [Griswold 02, Lustig 10, Uecker 14, Haldar 16, Kim 17], we propose a joint framework for reconstruction. We take undersampled measurements from all coils together as an input to a nonlinear reconstruction algorithm. The developed method is an extension of our optimization-based algorithm for single-coil reconstruction, which uses compound regularization (MRF prior + TV norm). The proposed method outperforms state-of-the-art CS-pMRI methods in terms of the quality and justifies the use of the proposed MRF based regularization in a multi-coil reconstruction scenario.

## 1.2 Publications resulting from this work

Article in international journals:

- Panić M., Aelterman J., Crnojević V., Pižurica A. 'Sparse recovery in magnetic resonance imaging with a Markov random field prior,' IEEE Transactions on Medical Imaging, Vol. 36, No. 10, pp. 2104-2115, Oct. 2017.

Articles in international conference proceedings:

- Panić M., Aelterman J., Crnojević V., Pižurica A. 'Compressed sensing in MRI with a Markov random field prior for spatial clustering of subband coefficients,' 24th European Signal Processing Conference (EUSIPCO), Budapest, Hungary, Aug. 29 - Sep. 02., pp.562-566, IEEE, 2016.

- Panić M., Vukobratović D., Crnojević V., Pižurica A. 'Greedy MRI reconstruction using Markov Random Field prior,' Proceedings of the IEICE Information and Communication Technology Forum (ICTF 2017), Poznań, Poland, Jul 4-6, Polish Association of Telecommunication Engineer, 2017.

- Panić M., Jakovetić, D., Crnojević, V. and Pižurica, A., 'Image Inpainting and Demosaicing via Total Variation and Markov Random Field-Based Modeling,' 26th Telecommunications Forum (TELFOR), Belgrade, Serbia, Nov. 26-27, pp. 1-4, IEEE, 2018.

- Panić M., Aelterman J., Crnojević V., Pižurica A. 'Multi-coil magnetic resonance imaging reconstruction with a Markov Random Field prior,' SPIE Medical Imaging, San Diego, USA, Feb. 19-21, Vol 10949, 2019. In print.

Conferences abstracts

- Panić M., Vukobratović D., Crnojević V., Pižurica A. 'Sparse MRI with a Markov Random Field Prior for the Subband Coefficients,' Third International Traveling Workshop on Interactions between Sparse models and Technology (iTWIST'16), Aalborg, Denmark, Aug. 24-26, pp. 56-58, 2016.

  Article in preparation for journal submission

- Panić M., Jakovetić, D., Vukobratović D., Crnojević V., Pižurica A. 'MRI reconstruction using Markov Random Field and Total Variation as composite prior,'.

## 1.3 Thesis organization

The thesis is organized as follows.

In Chapter 2, we present concisely the basic MRI principles starting from nuclear magnetic resonance physics towards the image acquisition process. For a better understanding of MR image creation, the process of MR signal detection is reviewed. Relaxation times T1 and T2 of net magnetization, respectively called spin-lattice and spin-spin, are introduced. An explanation of commonly used pulse sequence for signal detection named spin-echo sequence is given. The underlying image formation theory which refers to slice encoding and spatial localization is provided. We introduce a mathematical formulation of the MR image reconstruction problem which involves Fourier transform, and pay attention to the connection between the notions of a field of view (FOV) of MR image, its resolution and sampling rate in the $k$-space. We close the chapter with a brief overview of the traditional MRI reconstruction methods starting from the single-coil towards multi-coil acquisition procedures which do not include measurements undersampling.

Chapter 3 gives a short introduction to the CS theory and reviews the relevant state-of-the-art methods for the MR image recovery from partial data. We start by explaining the principal idea of sparse MRI given by Lustig et al. in their seminal work [Lustig 07]. They

explain how CS theory can help in reducing MRI acquisition time by keeping the same image quality after reconstruction. We then introduce fundamental notions used in CS theory, such as compressibility or sparseness as signal properties and restricted isometry property as the sensing procedure property. These properties need to be satisfied in order to guarantee the exact image reconstruction from undersampled data. Next, we present a review of standard CS reconstruction algorithms. Since MR images are compressible, we provide a description of important transformations for sparse image representation. We start by introducing the discrete wavelet transform (DWT), its non-decimated version (NDWT), and an efficient algorithme à trous [Holschneider 90] for the calculation of NDWT. Then we explain a non-decimated discrete shearlet transform which brings more efficient encoding of anisotropic features in images compared to the NDWT. We close the chapter with an overview of the state-of-the-art methods for CS-MRI reconstruction emphasizing those which concentrate on structured sparsity encoding and on those which will be fundamental for developing our methods in the following chapters.

Chapter 4 presents our optimization-based approach to CS-MRI with Markov Random Field (MRF) priors. We construct a signal support model based on an MRF prior and derive two different optimization-based CS-MRI reconstruction approaches. We start with a short introduction to the theory of MRF models, which are a special class of undirected probabilistic graphical models (PGMs). Then we explain the theorem which establishes the equivalence between the Gibbs distribution and MRF and provides a simple way of specifying the joint probability distribution. Using this property an MRF signal prior is proposed and together with the conditional model a final signal model is created. Based on the proposed signal model, we define new regularization functions which are utilized in our CS-MRI reconstruction algorithms. We derive our first optimization-based approach, with two concrete algorithms: LaSAL and LaSAL2, by incorporating the proposed regularization function into a constrained augmented Lagrangian (C-SALSA) iterative framework. The proposed regularization results in hard-thresholding of image coefficients based on the estimated signal support in the transform domain. We give an extensive performance and empirical convergence analysis of LaSAL and LaSAL2 and demonstrate their effectiveness in comparison with the state-of-the-art methods in the field. The results of this work are published in a conference paper [Panić 16a] and a journal paper [Panić 17a]. The potential of

the MRF-based regularization in combination with TV is further tested under another optimization framework called the fast composite splitting algorithm (FCSA). In the developed method, named FCSLaTV, we extend the proposed MRF model to its anisotropic version to achieve better model generalization and propose an efficient approach for automatic estimation of MRF parameters. We derive a new soft-thresholding regularization rule using the estimated signal support. Compared to LaSAL2, FCSLaTV has lower complexity which is reflected in the smaller number of parameters that must be initialized and in the smaller number of steps per one algorithm iteration. This is useful when the sampling rates are very low (below 25%) which leads to an unreliable estimation of the amount of noise present in the measurements. On the other side, when we have prior information about the noise variance or its reliable estimated value, the higher complexity of the LaSAL2 method is justified since it reveals a very subtle image structure. A detailed comparative analysis of the FCSLaTV algorithm is conducted with respect to the current state-of-the-art methods. This work is included in a journal submission in preparation.

In Chapter 5 we develop a greedy method for MR image recovery based on MRF context modeling and dubbed GreeLa to reflect greedy reconstruction with lattice regularization. GreeLa builds on a greedy reconstruction method LaMP (Lattice Matching Pursuit) of Cevher et al. in [Cevher 10b]. An important difference is that the LaMP was developed for the reconstruction of canonical sparse images from Gaussian random measurements. We extended this approach such that it is able to recover MRI images from undersampled Fourier coefficients. Moreover, compared to the LaMP we used different MRF model for the signal support and different model for the signal likelihood. First, we describe briefly the LaMP method of [Cevher 10b] and then we present in detail the proposed GreeLa method. After performing a gradient descent step and obtaining the temporary signal estimate, a proposed MRF based regularization is conducted in the transformation domain. We end the chapter with a comprehensive performance analysis on the same dataset that was used for evaluating the optimization-based methods. The results of this work are published in a conference paper [Panić 17b] and a conference abstract [Panić 16b].

In Chapter 6, we consider the problem of MR reconstruction in a parallel MRI (pMRI) setting using multi-coil data. We extend our approach, developed primarily for the single-coil data reconstruction, to the multi-coil problem scenario. Instead of separately reconstructing the

coil images and then obtaining the final MR composite image by combining them, we formulate a joint reconstruction framework, which takes as input measurements from all coils. In the original pMRI setting, acquisition time is reduced by introducing more coils that simultaneously acquire measurements that are related to different regions of the objects under scanning. Since all measurements from each coil are used, the reconstruction problem is well-posed. We consider a further reduction of acquisition time by involving the undersampling of coil measurements, which leads to an ill-posed reconstruction problem. The necessary regularization steps are then conducted on MR composite image using the previously developed compound regularization (MRF prior + TV norm). We give a comparison of the proposed method with the current state-of-the-art methods which consider CS-pMRI settings and analyze its performance on different datasets. The results of this work are published in the conference paper [Panić 19].

Chapter 7 concludes the thesis, provides discussions on the proposed methods and gives possible further research directions.

# 2

# Principles of MRI

*It was eerie. I saw myself in that machine.*
*I never thought my work would come to this.*
—Isidor Isaac Rabi

This chapter introduces the principles of magnetic resonance imaging (MRI) and reviews traditional methods for MRI image recovery. The principles of MRI formation lead to the mathematical formulation of the MR image reconstruction problem as a linear inverse Fourier problem or as an optimization problem. In the following, we will consider both formulations and review the relevant methods for MR recovery from measurements obtained using a single or multi-coil MRI acquisition scenario.

## 2.1 Basic principles of MRI

Magnetic resonance imaging (MRI) is an imaging modality used primarily in medical settings to deliver high-quality images of the interior of the human body. Based on the absorption and emission of energy in the radio frequency range of the electromagnetic spectrum, MRI is a non-ionizing technique well suited to image the non-bony parts or soft tissues of the body. MRI has its origin in nuclear magnetic resonance (NMR), a spectroscopic technique used by scientists to obtain microscopic chemical and physical information about molecules. The technique was called magnetic resonance imaging rather than nuclear magnetic resonance imaging (NMRI) because of the negative connotations associated with the word nuclear in the late 1970's [Hornak 06]. The resonance in the name refers to the equalization procedure between the frequency of an oscillating magnetic field and the 'precessional' fre-

quency of the spin of a nucleus (hence the 'nuclear') in an atom of a tissue.

Discovery of the proton spin nature was one of the initial sparks for the invention of the NMR concept. Isidor Isaac Rabi with his coworkers analyzed the interaction of the proton spin with a magnetic field in the 1930s and was the first who described the NMR phenomena. In NMR, the primary source of energy input is coming from a rotating magnetic field, discovered in 1882 by Nikola Tesla. Based on Rabi's work, Bloch and Purcell extended the quantum mechanical concepts to a measurement of an effect of the precession (change in the orientation of the rotational axis) of the hydrogen nucleus (the proton) spin around a magnetic field. The significant step from one-dimensional (1D) NMR signals towards two-dimensional (2D) images was made in 1973 with the seminal papers by Lauterbur and Mansfield [Lauterbur 73,Mansfield 77]. Knowing that the spin of the hydrogen nucleus in a magnetic field precesses about that field at the 'Larmor frequency', they introduced the idea that if a spatially varying magnetic field is introduced across the object, the Larmor frequencies are also spatially varying. This leads to a possibility of separating different frequency components of the signal giving the spatial information about the scanning object (spatial encoding of the data). Richard Ernst in 1975 connected the Fourier transform with the reconstruction of 2D MR images by using the switched magnetic field gradients in the time domain for spatial encoding.

In the following subsections, we review the basic principles of MRI. For a more detailed explanation, we refer the reader to [Wright 97, Kuperman 00, Brown 14, Ramos Llordén 18].

### 2.1.1 Interaction between proton spins and the Magnetic Field

The human body is primarily fat and water, where both have many hydrogen atoms. Approximately, 63% of our body is composed of hydrogen atoms. Every atomic nucleus possesses a nuclear magnetic moment that arises from the spin of the protons and neutrons which constitute the nucleus. When the human body is placed in the presence of a strong external magnetic field it becomes polarized in the sense that some proton magnetic moments coming from the hydrogen nucleus become aligned with that magnetic field. For these reasons, magnetic resonance imaging captures the NMR signal originating primarily from the hydrogen nuclei in water, fat and other organic molecules. Virtually all clinical MR

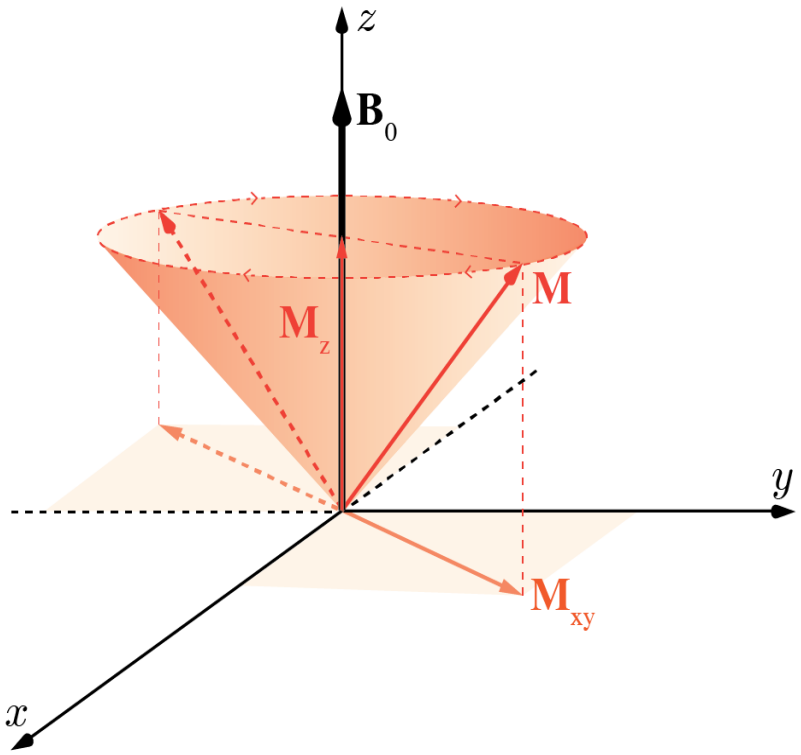



**Figure 2.1:** Precession of net nuclear magnetization vector $\mathbf{M}$ about $\mathbf{B}_0$ with angular (Larmor) frequency $\omega_L$

imaging relies on manipulating the nuclear magnetic field of hydrogen usually at a field strength in a range from 0.5 to 3.0 Tesla (T). The MRI image represents a digital map of the NMR signal intensity. To arrive at an MR image, the MRI scanner must create detectable NMR signals from tissues and spatially localize it.

Consider a volume of the human body that we would like to image using an MRI device. In that volume, there are many hydrogen nuclei (protons) each with an associated dipole magnetic moment $\mu$, which arises from the proton spins. A proton angular moment $I$ is colinear with the dipole magnetic moment $\mu$ and related to it through the following quantum mechanical relationship

$$\mu = \gamma I \tag{2.1}$$

where $\gamma$ is the gyromagnetic ratio of the nucleus. Without the presence of an external magnetic field, the direction of $\mu$ is completely random.

However, when we expose the hydrogen nuclei to a strong static external magnetic field the corresponding magnetic moments $\mu$ of proton spins tend to align themselves like a compass needle in the magnetic field of the Earth. The collective behavior of the spins of all protons captured in the observed volume is expressed through the macroscopic magnetization vector $\mathbf{M} = M_x\mathbf{i} + M_y\mathbf{j} + M_z\mathbf{k}$, where $\mathbf{i}, \mathbf{j}$ and $\mathbf{k}$ are unit vectors along the $x, y$ and $z$ axis, respectively. Therefore, the magnetization vector $\mathbf{M}$ is represented as the vector sum of the individual dipole moments $\mu$. Another name for this vector is the net nuclear magnetization vector. In the absence of an external magnetic field, the magnetization vector equals to zero ($\mathbf{M} = 0$) due to the random direction of each dipole magnetic moment.

Induction of an external static magnetic field $\mathbf{B}_0 = B_0\mathbf{k}$ in the observed volume, results in the precession of the proton spin about the field direction (see Fig. 2.1). The external field $\mathbf{B}_0$ forces the individual magnetic moments to align themselves in its direction. As a result, $M_z$ component is non-zero since it is a sum of the $z$-components of the individual dipole magnetic moments $\mu$. The transversal components of $\mathbf{M}$ $(M_x, M_y)$ are still zero since the individual spins have a random phase in $xy$-plane when they precess about the $z$-axis. The precession angular frequency for the $\mu$ is proportional to the strength of the external magnetic field $\mathbf{B}_0$

$$\omega_L = \gamma \|\mathbf{B}_0\|_2 = \gamma B_0 \tag{2.2}$$

and it is referred to as Larmor frequency, while (2.2) is referred to as the Larmor equation. This relation holds on the macroscopic level which is reflected in precession of $\mathbf{M}$ about $\mathbf{B}_0$ with Larmor frequency [Ramos Llordén 18]. Dynamics of change of magnetization vector $\mathbf{M}$, which represents evolution of an ensemble of spins in the presence of a magnetic field, can be described through celebrated Bloch equations [Bloch 46]. In more general form they describe the evolution of $\mathbf{M}$ along time $t$ under the presence of an arbitrary magnetic field $\mathbf{B} = B_x\mathbf{i} + B_y\mathbf{j} + B_z\mathbf{k}$:

$$\frac{\partial \mathbf{M}}{\partial t} = \underbrace{\gamma \mathbf{M} \times \mathbf{B}}_{\textbf{excitation}} - \underbrace{\begin{bmatrix} T_2^{-1} & 0 & 0 \\ 0 & T_2^{-1} & 0 \\ 0 & 0 & T_1^{-1} \end{bmatrix} \mathbf{M} + \mathbf{M}_0 \begin{bmatrix} 0 \\ 0 \\ T_1^{-1} \end{bmatrix}}_{\textbf{relaxation}} \tag{2.3}$$

where $\times$ represents the cross-product operation and $\mathbf{M}_0$ is the net nuclear magnetization vector at equilibrium. $T_2$ is the transverse relaxation


**Figure 2.2:** A 3 Tesla Skyra Magnetic Resonance Imaging (MRI) Scanner (SIEMENS). Image courtesy of Siemens-Healthcare.

time related to components $(M_x, M_y)$ of the magnetization vector $\mathbf{M}$ in the transversal plane $xy$, while $T_1$ is longitudinal relaxation time which refers to $M_z$ component along longitudinal axis $z$. (2.3) includes two fundamental processes of the NMR phenomenon, namely excitation process due to an applied magnetic field, and the relaxation process of vector $\mathbf{M}$ towards the system equilibrium.

In the next subsection, we describe the process of MR signal detection after the initiation of the dynamic magnetic field $\mathbf{B}_1$, i.e., a time-varying radiofrequency (RF) magnetic field. The $\mathbf{B}_1$ field is produced by driving electrical currents through specialized RF-transmit coils. For a better understanding of how the RF field $\mathbf{B}_1$ changes the motion of the individual nuclear magnetic moments and therefore the net magnetization vector $\mathbf{M}$, here we introduce a frame of reference concept. We will consider two frames of reference: one connected to the static magnetic field $\mathbf{B}_0$ called stationary or laboratory frame, depicted in Fig. 2.1, and other called rotating frame which is connected to the dynamic RF field $\mathbf{B}_1$ and precesses about the $z$-axis following the motion of the magnetization vector $\mathbf{M}$. In the following, we give a short description of the MRI device hardware and relate this description to the terms introduced in the previous paragraphs, which explain the essence of the MR image acquisition.

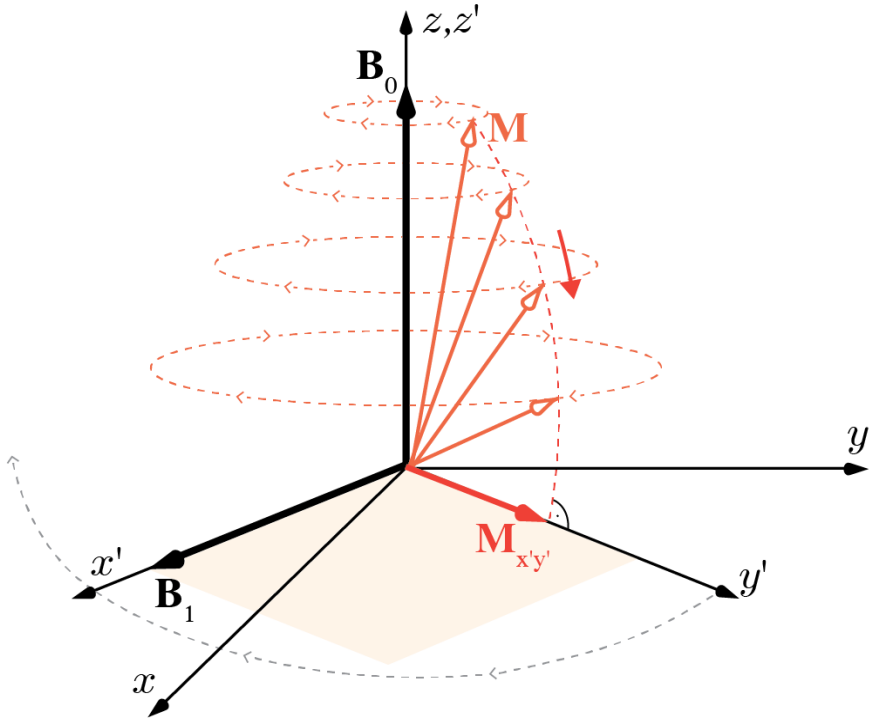



**Figure 2.3:** Motion of the net nuclear magnetization vector $\mathbf{M}$ in the presence of an RF field $\mathbf{B}_1$ in both frames of reference: laboratory $xyz$ and RF-rotating frame $x'y'z'$.

Fig. 2.2 shows a typical MRI scanner. The three main hardware components of the scanner are: a permanent superconducting magnet, an RF coil system and a gradient coil system. The superconducting magnet is an electromagnet made of superconducting wire and is responsible for the generation of the homogeneous static magnetic field $\mathbf{B}_0$. The usual strength of $\mathbf{B}_0$ is 1.5T or 3T in clinical MRI scanners, while in some scanners $\mathbf{B}_0$ has strength up to 11T [Nowogrodzki 18]. The main task of the superconducting magnet is to produce a homogeneous magnetic field over the whole region of interest. The RF system consists of two parts: the transmitter coil, used to generate the time-varying magnetic field $\mathbf{B}_1$, and the receiver coil, which converts the magnetization vector $\mathbf{M}$ into the MR signal $s(t)$. Today, MRI scanners contain multiple receiver coils, which are often collected into large arrays for use in parallel imaging applications. The magnetic field gradient system provides a time-varying magnetic field, which allows spatial encod-

ing procedure necessary for sampling in the so-called $\not{k}$-space where the spatial-frequency components of MR image are located. An important specification for the gradient system is the slew rate which is the maximum gradient strength divided by the rise time (time to achieve the maximum gradient strength). Most clinical scanners operate with slew rates around 200 $T/m/s$ [Ramos Llordén 18].

### 2.1.2 RF excitation and signal detection

Detection of the signal in the MR scanner is guided by the change of the magnetic flux in any nearby receiver coil. Based on Faraday's law of induction, the voltage (signal) induced in the coil depends linearly on the time change of the magnetic flux. The flux change comes as a consequence of the magnetization precession around the external magnetic field $\mathbf{B}_0$. To accomplish the magnetization precession, i.e., to stimulate the spin system, a transmit coil produces a time-varying excitation in the form of an RF magnetic field $\mathbf{B}_1$ with angular frequency denoted as $\omega_{RF}$ for a short time period (RF pulse) with the following characteristics: it is perpendicular to $\mathbf{B}_0$ and it rotates/oscillates at a resonance Larmor frequency. As a consequence of fulfillment, the resonance condition $\omega_{RF} = \omega_L$, the precession of $\mathbf{M}$ is continuously pushed away from the longitudinal direction $z$ (tilted by 90° in transversal plain Fig. 2.4). This leads to the so-called forced precession of the transversal components of $\mathbf{M}$ in the $xy$ plane which will induce an electromotive force in the receiver coil. Fig. 2.3 illustrates the forced precession in the laboratory frame $xyz$ (depicted also in Fig. 2.1) and the new RF-rotating frame of reference denoted with $x'y'z'$. For both frames the longitudinal direction $z = z'$ is equal. We see the movement of the magnetization vector $\mathbf{M}$ from the longitudinal direction $z$ under the influence of the RF magnetic field $\mathbf{B}_1$ presented in the RF-rotating frame of reference. An RF pulse that tilts the original longitudinal magnetization for an angle of 90° into the transverse, or $xy$ plane, is called a $\pi/2$-pulse. Using the complex notation, magnetization components in the transverse plane $xy$, often referred to as 'transverse magnetization', can be described by $M_{xy} = M_x + iM_y$ where $i = \sqrt{-1}$. In the following, we will use $M_{x'y'}$ to denote the transverse component of $\mathbf{M}$ in the RF-rotating frame of reference.

The magnetized spin system, perturbed by an RF pulse, will return to its equilibrium state and this process is called relaxation. Returning to the equlibrium state is characterizied by longitudinal $T_1$ and

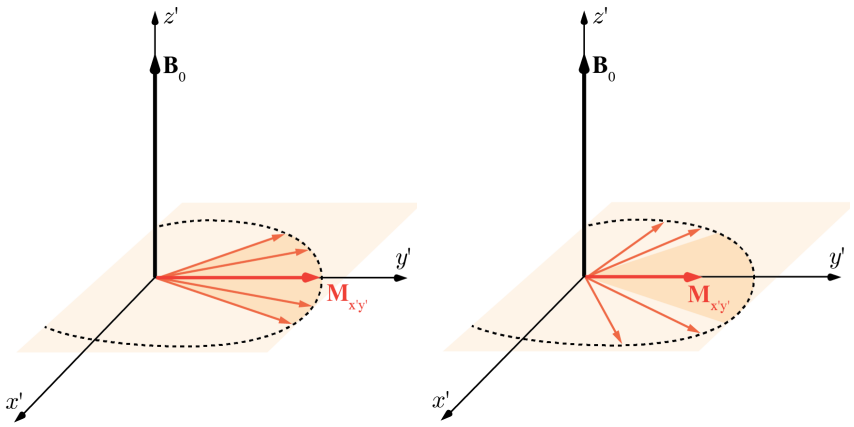



**Figure 2.4:** $T_2$ relaxation: The net nuclear magnetic dipoles associated with the spins dephase, causing progressive decrease in $M_{x'y'}$.

transverse $T_2$ relaxation times (see the second term in (2.3)). The longitudinal relaxation time $T_1$ refers to the relaxation process of the longitudinal component of magnetization $M_{z'} = M_z$, whereas the transverse relaxation time $T_2$ describes the relaxation curve of the transverse component $M_{x'y'}$. Dynamical behavior of the magnetization component $M_{z'}$ along the longitudinal $z'$ axis is explained through the following equation:

$$M_{z'}(t) = M_{z'}(0)e^{-\frac{t}{T_1}} + M_0(1 - e^{-\frac{t}{T_1}}) \tag{2.4}$$

where $M_0$ is the equilibrium value ($\mathbf{M}_0 = M_0\mathbf{k}$), while $M_{z'}(0)$ is the longitudinal component of the magnetization vector immediately after RF 90° pulse is applied. Similarly, the transverse component $M_{x'y'}$ of the magnetization vector has exponential behaviour in time characterized as:

$$M_{x'y'}(t) = M_{x'y'}(0)e^{-\frac{t}{T_2}} \tag{2.5}$$

where $M_{x'y'}(0)$ is the transverse component of the nuclear magnetization vector immediately after the RF 90° pulse. In Fig. 2.4, we show the process of spin dephasing caused by the interaction between spins of the same nuclear system, which results in a decrease of $M_{x'y'}$. The transverse relaxation (i.e., decay of $M_{x'y'}$), is usually much faster than longitudinal relaxation. Since the voltage (signal) in the reciever coil depends solely on the change of the transverse magnetization component, the detected signal known as free induction decay (FID) is rarely used alone since it quickly disappears. FID signals lay the foundations for more complex

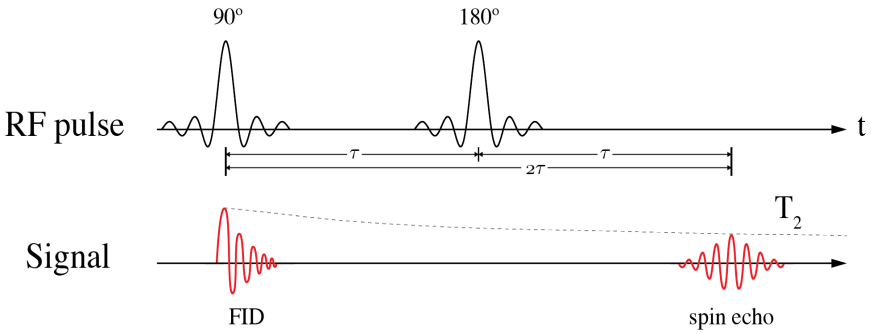



**Figure 2.5:** Spin echo sequence (figure adopted from [Ramos Llordén 18]). The 180° pulse flips the nuclear spins to the other side of the traverse plane (in -$y'$ direction). Spins are completely rephased after 2*$\tau$, thereby producing a measurable echo.

MR signals such as MR echoes. In the following, we review the technique invented by Hahn in 1950 [Hahn 50], which recovered to some extent the transverse magnetization loss.

A loss in phase coherence of the nuclear system spins results in an exponential decay of the transverse magnetization component $M_{x'y'}$ and therefore the FID signal. The technique of Hahn [Hahn 50] refocuses the spins that are progressively dephasing by inducing another RF pulse after the initial one and thereby creating a measurable echo which is known as a spin echo. Herman Carr and Edward Purcell pointed out the advantages of using a 180° pulse along $y'$-axis after the initial 90° pulse along the $x'$-axis. The resulting pulse sequence shown in Fig. 2.5 is called the spin-echo (SE) sequence. After application of the initial 90° pulse, fast spins start to precess clockwise about the $z$-axis faster than the slow spins, resulting in losing phase coherence (see Fig. 2.4). The application of the 180° pulse after some period of time $\tau$ rotates the magnetization vector around the $x'$-axis, hence, flipping the nuclear spins by 180° in -$y'$-axis direction or to the other side of the transverse plane. Continuing the precession of the spins clockwise, now slow spins lead in front of the fast ones. Progressively, fast spins catch slow spins and hence complete refocusing is achieved at a time 2*$\tau$ called echo time when the spin echo is produced.

The time varying (decreasing) tranverse component $M_{x'y'}(t)$ of the magnetization vector induce a voltage in a reciver coil. The $M_{x'y'}(t)$ can be related to the transverse component $M_{xy}(t)$ in a laboratory frame through the following equation $M_{xy}(t) = M_{x'y'}(t)e^{-i\omega_{RF}t}$. The signal of

interest in MRI, denoted with $s(t)$, is represented through the complex envelope of the signal of induced voltage. The complex signal $s(t)$ is the integrated product of so-called reciever coil sensitivity $B_{xy}^R$ and magnitude of $M_{xy}(t)$ (i.e. $M_{x'y'}(t)$), in transverse plane $xy$ over the volume $\Omega$ enclosed by that coil:

$$s(t) = \int_{\Omega} B_{xy}^R(\mathbf{r}) |M_{xy}(\mathbf{r}, t)| d\mathbf{r} \tag{2.6}$$

where $\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$.

From (2.6) we see that all spin systems in the object contribute forming the signal $s(t)$, which means that it is impossible to determine $M_{xy}(\mathbf{r}, t)$ at every point $\mathbf{r}$ in the scanned object using only information from $s(t)$. In the following subsection, we introduce the spatial localization or the encoding concept which is necessary for MR image formation.

### 2.1.3 MR image formation

In order to resolve the spatial contribution of the magnetization vector inside the object, a spatially selective excitation over a restricted region must be conducted. This will contribute to the encoding of the signal spatial location during the data acquisition procedure.

The purpose of spatially selective excitation is to direct magnetization in a thin spatial slice or section of an object under scanning, say of thickness $\Delta z$ along the $z$ axis, into the transverse plane. This is accomplished in two steps. First, a linear magnetic gradient field $G_z$ that varies along the $z$-direction is introduced in the system. This causes the Larmor frequency to vary linearly in the $z$-direction:

$$\omega_L = \gamma(\|\mathbf{B}_0\|_2 + G_z z). \tag{2.7}$$

Then, as a second step, a slice-selective RF pulse with a significant signal energy only over a limited range of temporal frequencies $\Delta f$ is applied. This range corresponds to the Larmor frequencies in the slice and determines the $\Delta z$ through the following relation $\Delta z = 2\pi\Delta f/\gamma G_z$. Therefore if the spin systems located at $|z - z_0| \leq \Delta z$ are excited, then the resonance frequency of the RF pulse is:

$$\omega_{RF} = \gamma(\|\mathbf{B}_0\|_2 + G_z z_0) \tag{2.8}$$

Omiting for simplicity practical considerations such as the effect of the slice profile, the MRI signal $s(t)$ can be interpreted as a 2D Fourier transform of the image obtained by averaging slices $\rho(x, y, z, t)$ that represent physical densities of three-dimensional object within $|z - z_0| \leq \Delta z$.

With an isolated volume using selective excitation, an image of a planar portion of the three-dimensional object can be generated by manipulating the phase of $M_{xy}(\mathbf{r},t) = |M_{xy}(\mathbf{r},t)|e^{-i\omega_{RF}t}$ [Wright 97]. Introducing a magnetic gradient field $\mathbf{B_G} = B_{G_x}\mathbf{i} + B_{G_y}\mathbf{j} + B_{G_z}\mathbf{k}$ along the $z$-direction, a spin system located at point $\mathbf{r}$ changes the frequency as follows:

$$\omega(\mathbf{r}) = \omega_L + \gamma B_{G_z}(\mathbf{r}). \tag{2.9}$$

The $z$-component of the magnetic gradient field $B_{G_z}$ varies linearly with $x, y$ and $z$:

$$B_{G_z}(\mathbf{r}) = G_x x + G_y y + G_z z = \langle \mathbf{G}, \mathbf{r} \rangle \tag{2.10}$$

where $\mathbf{G} = G_x\mathbf{i} + G_y\mathbf{j} + G_z\mathbf{k}$ is the gradient vector. This results in phase changing of the transverse magnetization component $M_{xy}(\mathbf{r},t) = |M_{xy}(\mathbf{r},t)|e^{-i\omega_{RF}t}e^{-i\gamma\langle \mathbf{G},\mathbf{r}\rangle t}$ and the MR signal transforms into:

$$s(t) = \int_\Omega B_{xy}^R(\mathbf{r})|M_{xy}(\mathbf{r},t)|e^{-i\gamma\langle \mathbf{G},\mathbf{r}\rangle t}d\mathbf{r}. \tag{2.11}$$

If we consider the gradient fields which change with time, a more general formulation can be given:

$$s(t) = \int_\Omega B_{xy}^R(\mathbf{r})|M_{xy}(\mathbf{r},t)|e^{-i\gamma\int_0^t\langle \mathbf{G}(t'),\mathbf{r}\rangle dt'}d\mathbf{r}. \tag{2.12}$$

The MR signal provides information about the spatial frequency content of the image instead of directly about the spatial positioning of information in the image. Therefore the MR signal is acquired in spatial frequency domain of the image, so called $k$-space, whose coverage depends on gradient field. Denoting the trajectories in the $k$-space by $\boldsymbol{\kappa}(t) = k_x(t)\mathbf{i} + k_y(t)\mathbf{j} + k_z(t)\mathbf{k}$ and letting:

$$\boldsymbol{\kappa}(t) = \frac{\gamma}{2\pi}\int_0^t \mathbf{G}(t')dt' \tag{2.13}$$

the equation (2.12) can be reformulated as:

$$s(t) = \int_\Omega B_{xy}^R(\mathbf{r})|M_{xy}(\mathbf{r},t)|e^{-2\pi i\langle \boldsymbol{\kappa}(t),\mathbf{r}\rangle}d\mathbf{r} \tag{2.14}$$

With the selective excitation procedure a slice in the three-dimensional object is considered for imaging. This constraints the source of the MR

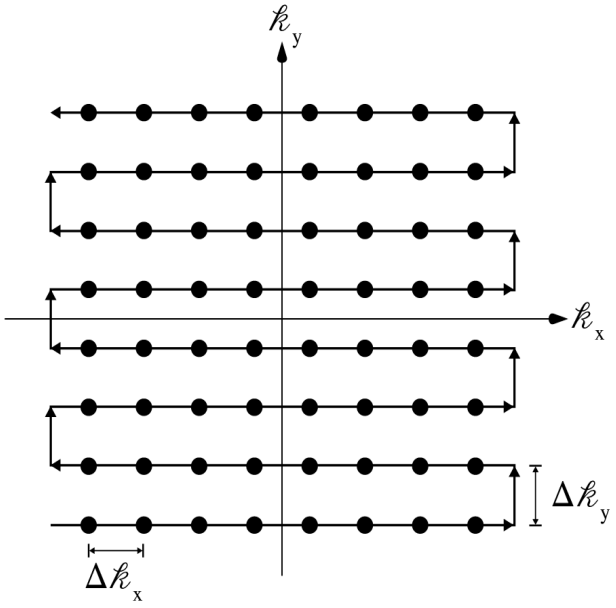



**Figure 2.6:** The $\kappa$-space coverage for a 2D example. Each dot represents a sampled point. Lines of connected dots are shown to be along the read direction, referring to data collected during the same read period. The arrows indicate the chronological order of data acquisition.

signal $s(t)$ coming from spin system in $xy$-plane and the expression from (2.14) becomes

$$s(t) = \int_x \int_y B_{xy}^R(x,y)|M_{xy}(x,y,t)|e^{-2\pi i(\kappa_x(t)x+\kappa_y(t)y)}dxdy \qquad (2.15)$$

where the $s(t)$ is expressed in terms of only $x$ and $y$ components of the magnetization vector and the corresponding components of $\kappa$-space trajectories. For a fixed acquisition time point $t_{\text{acq}}$, $s(t_{\text{acq}})$ is the 2D Fourier transform of the image which refers to the planar portion of the object under scanning $\rho(x,y) = B_{xy}^R(x,y)|M_{xy}(x,y)|$ evaluated at frequency point $(\kappa_x(t_{\text{acq}}), \kappa_y(t_{\text{acq}}))$, where we assume that $|M_{xy}(x,y)|$ is constant during data acquisition $t \in [0, t_{\text{acq}}]$ [Wright 97]. The set of complex values $s(t_p)$ for every $p \in \mathbb{Z}$, where $\mathbb{Z}$ denotes the set of integer numbers, represent the $\kappa$-space data while $(\kappa_x(t_p), \kappa_y(t_p))$, $\forall p \in \mathbb{Z}$ is the set of $\kappa$-space points where the data is acquired.

In general, sampling the $\kappa$-space at points $(\kappa_x(t), \kappa_y(t))$ cannot

be completed after a single excitation due to physical limitations (finite relaxation time of the dipoles) and technical limitations (slew rate limits on $G_x(t)$ and $G_y(t)$). Thus, the sampling is performed using a sequence of $n$ excitation-acquisition cycles with a determined time of repetition (TR). The usual sampling method is along a line in $\not{k}$-space, which corresponds to $\not{k}_y = const$. The procedure is shown in Fig. 2.6 where for $n$ lines, imaging time is $n \times TR$. According to the Fourier transform theory, the $\not{k}$-space sampling pattern affects the resolution and the Field of View (FOV) of the formed MR image. The sampling intervals $\Delta\not{k}_y$ and $\Delta\not{k}_x$ determine the FOV in $y$ and $x$ dimensions as follows:

$$FOV_y = \frac{1}{\Delta\not{k}_y}, \quad FOV_x = \frac{1}{\Delta\not{k}_x}. \tag{2.16}$$

From the Nyquist sampling theorem, we know that in order to avoid aliasing artifacts in the reconstructed image, $FOV_y(FOV_x)$ should be greater than the extent of the object in $y(x)$ dimension. Image resolution depends on the range of spatial frequencies sampled in both dimensions. Let's say if we constrain the $\not{k}$-space by a rectangular region of dimensions $W_{\not{k}_x} \times W_{\not{k}_y}$, then after we applied the inverse Fourier transform, the obtained image will have resolution $\delta_x = \frac{1}{W_{\not{k}_x}}$ and $\delta_y = \frac{1}{W_{\not{k}_y}}$, where $\delta_x$ and $\delta_y$ are the width and height of the pixel. Also the FOV and image resolution are connected through the following relation:

$$\delta_x = \frac{FOV_x}{N_x}, \quad \delta_y = \frac{FOV_y}{N_y} \tag{2.17}$$

where $N_x$ and $N_y$ are the number of pixels in $x$ and $y$ dimension, respectively. With the adopted strategy of sampling along the $\not{k}_x$ line for $\not{k}_y = const$, $FOV_x$ can always be increased by reducing the sampling interval. On other hand, the lines in $\not{k}_y$ directions are acquired at time intervals equal to TR which demands a trade-off between $FOV_y$, $\delta_y$ and imaging time in this dimension.

The relation given in (2.15) is of huge importance since it allows to interpret the formation of an MR image as an inverse Fourier problem. In the ideal case, if we know all the values of the Fourier transform of an image $\rho(x, y)$ in the acquired $\not{k}$-space data set, the image $\rho(x, y)$ can be analytically calculated using the inverse Fourier transform. However, in real life, there are a lot of challenges that we encounter in order to solve the inverse Fourier problem and form the MR image. One of these is the trade-off between the imaging time and the quality of the reconstructed image. In this thesis, we develop signal processing algorithms based on

a theory of compressed sensing (CS), which defines conditions when the signal reconstruction is possible if we sample below the Nyquist sampling rate. With this, the scanning time can be reduced, meeting the compromise between the acquisition time and the desired image resolution.

In practice, most of the MR scanners have more than one RF detector coil, which acquires simultaneously portions of the MR signal. The technique that uses smaller coils and combines or feds them into separate receiver chains is called 'phased array' and comes from the antenna theory. This technology, introduced in the pioneering work of Roemer [Roemer 90], was conceived with the purpose of increasing the signal-to-noise ratio (SNR) of the reconstructed MR image.

In parallel imaging methods in MRI (pMRI) [Blaimer 04, Pruessmann 06,Pruessmann 99,Lustig 10,Lai 10], which have been used frequently in the last decade, differential weighting of signals from multiple small surface coils are used to determine the spatial origin of the signal, thus reducing the need for time-consuming gradient-encoding steps explained previously. These small-diameter surface coils are positioned near the patient and have high sensitivity but limited anatomical coverage. By combining them into large arrays it is possible to obtain high signal-to-noise ratio and large fields of view. In order to get proper spatial encoding, the individual coil elements should be free of magnetic interactions. This requires that the large overlap between the coils, which has been seen in most phased-array configurations, is generally avoided for parallel imaging. Therefore, while all parallel imaging coils are 'phased arrays', not all phased array coils are suitable for parallel imaging. The coil segments are generally produced in a greater number than receiver channels because the coil elements are cheap to produce, whereas receiver channel technology is expensive, requiring a complete chain of amplifiers, digitizing circuitry, and computational engine to process the individual signals collected.

If the coil sensitivities are known, the MR image reconstruction problem in a phased-array or multi-coil system can be posed as determination of $|M_{xy}(\mathbf{r},t)|$ given the measurements:

$$s_l(t) = \int_\Omega B^R_{xyl}(\mathbf{r})|M_{xy}(\mathbf{r},t)|e^{-2\pi i\langle\boldsymbol{\kappa}(t),\mathbf{r}\rangle}d\mathbf{r}. \tag{2.18}$$

with transversal coil sensitivity profile denoted as $B^R_{xyl}(\mathbf{r})$ and $l = 1,2,3,...,L$ where $L$ is the number of coils in an array. In the following, we review standard MR reconstruction algorithms, both for single-coil and multi-coil acquisition systems.

## 2.2 MRI reconstruction algorithms

### 2.2.1 Linear inverse Fourier problem

Let us assume that the measurements $s(t_p)$ with $\forall p \in \mathbb{Z}$, are acquired within a time interval around $t_1$ i.e. $t_0 \leq t_1 \leq t_2$ where the temporal change of transverse component of magnetization vector $M_{xy}(x, y, t_1)$ is negligible and therefore we can omit $t_1$ as an argument. This type of acquisition is called a single-shot acquisition and therefore we can perfectly assume that all $\mathscr{k}$-space points correspond to the same continuous image $\rho(x, y) = B^R_{xy}(x, y)|M_{xy}(x, y)|$. Furthermore, we can simplify the notation by denoting the $\mathscr{k}$-space measurements $s(t_p)$ with $s(\boldsymbol{\kappa}_p)$ where $\boldsymbol{\kappa}_p$ represent the points in $\mathscr{k}$-space where the trajectory is evaluated at $t = t_p$, $p \in \mathbb{Z}$. With this simplification the expression in (2.14) can be reformulated as:

$$s(\boldsymbol{\kappa}_p) = \int_{\Omega_{xy}} \rho(\mathbf{r}_{xy}) e^{-2\pi i \langle \boldsymbol{\kappa}_p, \mathbf{r}_{xy} \rangle} d\mathbf{r}_{xy}, \quad \forall p \in \mathbb{Z} \tag{2.19}$$

where $\mathbf{r}_{xy} = r_x \mathbf{i} + r_y \mathbf{j}$ represents the vector of coordinates in the planar subspace $\Omega_{xy}$. The $\rho(\mathbf{r}_{xy}) = B^R_{xy}(\mathbf{r}_{xy})|M_{xy}(\mathbf{r}_{xy})|$ represents our continuous image of the planar portion of the three-dimensional object, which we would like to reconstruct. Further, simplifying the notation of $\mathbf{r}_{xy}, \Omega_{xy}$ with $\mathbf{r}, \Omega$ and using the approach from [Fessler 10] we can approximate the physical density of the object under scanning $\rho(\mathbf{r})$ with a finite series expansion as:

$$\rho(\mathbf{r}_{xy}) = \sum_{n=1}^{N} x_n b(\mathbf{r} - \mathbf{r}_n) \tag{2.20}$$

where $b(\cdot)$ denotes the object basis function, $\mathbf{r}_n$ is the center of the $n^{th}$ translated basis function and $N$ is the number of parameters. Here we denote with $x_n$ a discrete samples of the continuous-space function $\rho(\mathbf{r})$. For simplicity, hereafter we use rect basis functions $b(\mathbf{r}) = \text{rect}(\mathbf{r}/w)$, i.e., square pixels of dimension $w$, so $N$ is the number of pixels. Many other possible basis function choices can be considered, all of which are imperfect because the true object never satisfies the parametric model in (2.20) exactly. Nevertheless simple basis functions can provide useful approximations. Substituting the equation (2.20) in equation (2.19) we arrive at:

$$s(\boldsymbol{\kappa}_p) = \sum_{n=1}^{N} a_{pn} x_n, \quad \forall p \in \mathbb{Z} \tag{2.21}$$

with

$$a_{pn} = \int_{\Omega} b(\mathbf{r} - \mathbf{r}_n) e^{-2\pi i \langle \boldsymbol{\kappa}_p, \mathbf{r} \rangle} d\mathbf{r}, \quad \forall p \in \mathbb{Z}, \quad n = 1, ..., N. \tag{2.22}$$

With the fact that the basis function are highly localized in practice, the following approximations are nearly always used:

$$a_{pn} \approx e^{-2\pi i \langle \boldsymbol{\kappa}_p, \mathbf{r}_n \rangle} \quad \forall p \in \mathbb{Z}, \quad n = 1, ..., N. \tag{2.23}$$

which means that subintegral function from (2.22) is different from zero only when $\mathbf{r} = \mathbf{r}_n$. We next introduce a vector $\mathbf{x} = (x_1, x_2, x_3, ..., x_N)$ which represents, in a rasterized form, the discretized image of the continuous object $\rho(\mathbf{r})$. Knowing that the number of sampling points in $\mathcal{k}$-space is not indefinite in practice but equal to some number $M$, i.e. $p \in [0, M]$, we can then define $\mathbf{y} = (s(\boldsymbol{\kappa}_1), s(\boldsymbol{\kappa}_2), s(\boldsymbol{\kappa}_3), ..., s(\boldsymbol{\kappa}_M))^T$ as a measurement vector, and $\mathbf{A} = \{a_{m,n}\} \in \mathbb{C}^{M \times N}$ as the so-called Fourier encoding matrix with entries:

$$a_{m,n} = e^{-2\pi i \langle \boldsymbol{\kappa}_m, \mathbf{r}_n \rangle}. \tag{2.24}$$

Then the MR image reconstruction problem (without noise) can be posed as an algebraic linear inverse problem in the following matrix-vector form

$$\mathbf{y} = \mathbf{A}\mathbf{x}. \tag{2.25}$$

Assume that a total number of $\mathcal{k}$-space data points, denoted previously by $M$, are acquired fulfilling the Nyquist conditions on a Cartesian grid. Further, if $M = N$, where $N$ is the number of $\mathbf{r}_n$ samples in the $FOV$ of discretized image $\mathbf{x}$, then $\mathbf{A}$ is nothing more than a (scaled) DFT matrix. In this particular case, the recovery of image $\mathbf{x}$ from acquired measurements $\mathbf{y}$ can be calculated using the discrete inverse fast Fourier transform (IFFT). However, when a non-Cartesian $\mathcal{k}$-space schemes are used this is no longer possible and then an inverse version of a nonuniform FFT (NUFFT) is employed [Fessler 03]. One of the advantage of using non-Cartesian $\mathcal{k}$-space trajectories over the Cartesian is efficient use of MR gradient hardware and therefore rapid coverage of $\mathcal{k}$-space [Wright 14]. Additionally, many non-Cartesian trajectories are much more tolerant to under-sampling [Peters 06]. For radial $\mathcal{k}$-space sampling, reducing the number of radial profiles has a more clear effect in the high-frequency area rather than in the low-frequency region, where most of the energy is found. The resulting artifacts, which come in form of streaking artifacts, are shown to be incoherent [Ramos Llordén 18]. In this situation or with a presence of noise in the MRI measurements, the image reconstruction problem can be considered as an optimization problem which will be discussed in the following.

### 2.2.2 MR reconstruction from single coil measurements

A more realistic MRI acquisition model is usually considered in literature [Fessler 10] where the presence of white complex Gaussian noise $\mathbf{n} \in \mathbb{C}^N$ is included in measurements:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}. \tag{2.26}$$

Without involving any additional information about image $\mathbf{x}$, reconstruction is made by solving the linear least squares (LLS) optimization problem:

$$\min_{\mathbf{x}\in\mathbb{C}^N} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2. \tag{2.27}$$

Since $\mathbf{n}$ is Gaussian noise, the formulation given in (2.27) is equivalent to the maximum likelihood (ML) estimator of the image [Den Dekker 14]. The ML estimator can be derived analytically as:

$$\hat{\mathbf{x}} = \left(\mathbf{A}^H\mathbf{A}\right)^{-1}\mathbf{A}^H\mathbf{y} \tag{2.28}$$

if $(\mathbf{A}^H\mathbf{A})$ is invertible, where $\mathbf{A}^H$ is the Hermitian matrix of $\mathbf{A}$ and $\hat{\mathbf{x}}$ represents estimated (reconstructed) image. In practice storing the matrix $\mathbf{A}$ explicitly is inefficient. Therefore multiplication with a matrices $\mathbf{A}$ and $\mathbf{A}^H$ are implemented with a fast Fourier transform (FFT) and its inverse version (IFFT) respectively. The MR reconstruction problem often involves additional constraints on $\mathbf{x}$ leading to more complex cost functions than $\ell_2$ norm. One of the constraints imposed on $\mathbf{x}$ is the so-called phase-constraint formulation [McGibney 93, Samsonov 04, Bydder 05, Samsonov 10, Blaimer 16]. We can write the complex valued image $\mathbf{x}$ as:

$$\mathbf{x} = \mathbf{\Phi}\mathbf{x}_+ \tag{2.29}$$

where $\mathbf{x}_+ \in \mathbb{R}_+^N$ is the magnitude image and $\mathbf{\Phi} = \mathrm{diag}(e^{i\phi_\mathbf{x}})$ is a diagonal matrix with entries of the vector $e^{i\phi_\mathbf{x}}$ along the diagonal, where $\phi_\mathbf{x} \in \mathbb{R}^N$ is the phase of $\mathbf{x}$. Then we can define the phase-constraint MR reconstruction problem as:

$$\min_{\mathbf{x}_+\in\mathbb{R}_+^N} \|\mathbf{y} - \tilde{\mathbf{A}}\mathbf{x}_+\|_2^2. \tag{2.30}$$

where $\tilde{\mathbf{A}} = \mathbf{A}\mathbf{\Phi}$. Once the magnitude image $\mathbf{x}_+$ is recovered with (2.28), the estimate of the phase $\phi_\mathbf{x}$ can be obtained from very low-resolution fully sampled $\mathscr{k}$-space information [Lustig 07], since the phase image varies slowly compared to the magnitude image.

When a non-uniform Cartesian grid is used for sampling and generating of $\mathbf{y}$, additional processing of the measurements data is needed since the IFFT can not be used straightforwardly. The coordinates, where the non-Cartesian data points are sampled on, are besides integers also and rational numbers. They do not necessarily correspond with the implicit coordinates that the IFFT algorithm expects the data points to be sampled on [Aelterman 14a]. This issue is solved by interpolating the known non-Cartesian data in the Cartesian coordinate system that the IFFT uses, allowing in that way utilization of the IFFT algorithm in reconstruction. The combination of interpolation and IFFT algorithm is called the inverse non-uniform fast Fourier transform (INUFFT) [Fessler 03]. Since interpolation and IFFT are linear transformations then we can express INUFFT and its adjoint version, i.e., 'forward' NUFFT as a linear operators in the following matrix forms $\mathbf{F}^H$ and $\mathbf{F}$ respectively. Note that $\mathbf{F}^H$ and $\mathbf{F}$ are not rectangular matrices because the number of non-uniformly sampled points can be different from the number of points in the Cartesian grid. Direct application of an ML estimator from (2.28) for image reconstruction now with operators $\mathbf{F}^H$ and $\mathbf{F}$ is computationally too demanding. Instead, a so-called regridding technique is proposed [Johnson 09, Rasche 99]. Regridding, also called density compensation, uses a diagonal regridding weight matrix $\mathbf{W}$ to rescale the $\mathscr{k}$-space data

$$\hat{\mathbf{x}} = \mathbf{F}^H \mathbf{W} \mathbf{y} \tag{2.31}$$

prior to apply the adjoint NUFFT to estimate image $\hat{\mathbf{x}}$. Using regridding is a crude approximation of the ML estimator from (2.27) which together with the interpolation involved in INUFFT leads to the presence of artifacts in the reconstructed image $\hat{\mathbf{x}}$. In order to reduce image artifacts the problem formulation in (2.27), or its phase-constrained counterpart in (2.30), are reformulated with addition of the regularization term $R(\mathbf{x})$, which incorporate some prior knowledge about the type of image we expect to recover, as

$$\min_{\mathbf{x} \in \mathbb{C}^N} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda R(\mathbf{x}) \tag{2.32}$$

where $\lambda$ is the regularization parameter. Involvement of a regularization term in objective function is also necessary for the case when the sensing matrix is not well conditioned and the unregularized LLS solution can lead to undesirable noise amplification [Fessler 10]. The usual choice for the regularization term is Tihonov regularization $R(\mathbf{x}) = \|\mathbf{x} - \mathbf{x}_0\|_2^2$ where a reference image $\mathbf{x}_0$, which can be available from a previous MR scan or can be estimated iteratively, incorporates prior knowledge about desired

solution. A disadvantage of this choice is a biased estimate towards the reference image, which in particular for the zero reference image, forces all pixel values in the estimated $\hat{\mathbf{x}}$ towards zero, thus reducing the overall image contrast [Fessler 10]. Another common regularization choice is total variation (TV) norm defined in original isotropic form

$$R^{\mathrm{i}}_{\mathrm{TV}}(\mathbf{x}) = \sum_i \sum_j \sqrt{|\mathbf{x}_{i+1,j} - \mathbf{x}_{i,j}|^2 + |\mathbf{x}_{i,j+1} - \mathbf{x}_{i,j}|^2}, \tag{2.33}$$

or in anisotropic form

$$R^{\mathrm{a}}_{\mathrm{TV}}(\mathbf{x}) = \sum_i \sum_j |\mathbf{x}_{i+1,j} - \mathbf{x}_{i,j}| + |\mathbf{x}_{i,j+1} - \mathbf{x}_{i,j}|, \tag{2.34}$$

where indices $i$ and $j$ denotes rows and columns of the image $\mathbf{x}$ respectively. In literature the following definition of anisotropic form of TV norm can also be found: $R^{\mathrm{a}}_{\mathrm{TV}} = (\mathbf{x})\|\mathbf{D}\mathbf{x}\|_1$ where $\mathbf{D}$ is a finite difference matric. In this thesis we will focus on isotropic form of TV norm denoted in the following chapters as $\|\mathbf{x}\|_{\mathrm{TV}} = R^{\mathrm{i}}_{\mathrm{TV}}(\mathbf{x})$. This type of regularization biases the reconstructed image towards a piece-wise constant image with the appealing property that edges are sharply defined while the noise-type variations are substantially removed. However, involving TV regularization during reconstruction can lead to MR images with the undesirable appearance of cartoon-type images, also known, as staircase effect [Fessler 10]. For the minimization in (2.32) an iterative algorithms are needed. The algorithm construction greatly depends on choice of regularization term $R(\mathbf{x})$. For differentiable chosen regularizers the conjugate gradient algorithm is a natural choice [Fessler 10, Pruessmann 01, Sutton 03]. For nondifferentiable regularizers like in (2.34), more sophisticated algorithms are needed and this is an active research area [Ramani 10, Goldstein 09].

The problem formulation given in (2.32) is also used when the measurements are sampled below the Nyquist sampling rate. In the following chapter, we give a detailed overview of the methods which solve the problem in (2.32) in this case.

### 2.2.3 Parallel MRI (pMRI) methods

In the mid-nineties, great progress in the development of pMRI methods had taken place, thereby producing plenty of different and somewhat related parallel imaging reconstruction techniques and strategies [Blaimer 04]. Currently, the most well-known methods are sensitivity

encoding (SENSE) [Pruessmann 99] and generalized autocalibrating partially parallel acquisitions (GRAPPA) [Griswold 02].

### 2.2.4 SENSE method

The SENSE pMRI reconstruction method [Pruessmann 99, Blaimer 04] can be characterized as an image domain reconstruction method since it combines reduced $FOV$ coil images to reproduce the full composite image. Reduced $FOV$ coil images are obtained with accelerated acquisition by increasing the distance between the $k$-space trajectory lines (see Fig. 2.6) in the $k_y$ direction by a factor $N_p$. Then each pixel in the individual reduced $FOV$ coil image will contain information from multiple $N_p$ equidistantly distributed pixels in the desired full $FOV$ image. The undersamping in $k_y$ direction results in aliasing artifacts along direction of $y$-axis in each coil image (see Fig 3.1). These $N_p$ pixels will be weighted by the coil sensitivity $C$ at the corresponding location in the full $FOV$. Therefore, knowing the coil sensitivities allows pixels that are aliased to be separated by means of linear algebra. If we introduce the index $l$ which counts the number of coils from 1 to $N_c$, then the signal value in one pixel of $l$-th coil image $m_l$ at location $(x, y)$ can be written as:

$$m_l(x,y) = C_l(x,y_1)\mathbf{x}(x,y_1) + .. + C_l(x,y_{N_p})\mathbf{x}(x,y_{N_p}) \tag{2.35}$$

where $C_l$ is the sensitivity of the $l$-th coil, and $y_1, y_2, ..., y_{N_p}$ represent positions for $N_p$ pixels of the full $FOV$ image $\mathbf{x}$ in direction of $y$-axis, which contribute to the value of $m_l(x, y)$. Introducing the index $k = 1, 2..., N_p$, which specify the locations of the pixels involved, the equation (2.35) can be rewritten as

$$m_l = \sum_{k=1}^{N_p} C_{l,k}\mathbf{x}_k. \tag{2.36}$$

With all coils, a set of $N_c$ linear equations with $N_p$ unknowns can be established and represented using matrix notation

$$\mathbf{m} = \mathbf{C}\mathbf{x}_{N_p} \tag{2.37}$$

where matrix $\mathbf{C} \in \mathbb{C}^{N_c \times N_p}$ contains estimated sensitivities for each coil and all $N_p$ positions. The vector $\mathbf{x}_{N_p}$ lists the $N_p$ pixels in the full $FOV$ image and the vector $\mathbf{m}$ contains the complex values at the chosen pixel position from all coil images. Knowing the estimated complex

sensitivities at the corresponding positions, $\mathbf{x}_{N_p}$ can be calculated using a generalized inverse of the sensitivity matrix $\mathbf{C}$:

$$\mathbf{x}_{N_p} = (\mathbf{C}^H\mathbf{C})^{-1}\mathbf{C}^H\mathbf{m}. \tag{2.38}$$

To simplify matters, the issue of noise correlation is not addressed in (2.38). However, to account for levels and correlations of stochastic noise in the received data, which can be especially important when the receiver coils are not completely decoupled, a receiver noise matrix $\mathbf{R_n}$ with dimension $N_c \times N_c$ is included in (2.38) in the following way:

$$\mathbf{x}_{N_p} = (\mathbf{C}^H\mathbf{R}_\mathbf{n}^{-1}\mathbf{C})^{-1}\mathbf{C}^H\mathbf{R}_\mathbf{n}^{-1}\mathbf{m}. \tag{2.39}$$

A detailed description is given in [Pruessmann 99]. The estimator given in (2.38) for $N_p$ pixels in the full $FOV$ image can be used if the the matrix inversion $(\mathbf{C}^H\mathbf{C})^{-1}$ can be performed. This is possible if the number of pixels to be separated $N_p$ in full $FOV$ image doesn't exceed the number of coils $N_c$ in the receiver array. The SENSE algorithm repeats estimator from (2.38) or (2.39) for every pixel location in the reduced $FOV$ image to finally reconstruct the full $FOV$ image satisfying condition $N_p \leq N_c$. An important issue with SENSE is in the accurate estimation of the coil sensitivity matrix $\mathbf{C}$, which can be alleviated with the GRAPPA method explained next.

### 2.2.5 GRAPPA method

The GRAPPA method [Griswold 02, Blaimer 04], unlike SENSE, does not require knowledge of the coil sensitivities and reconstructs the image purely in the $k$-space domain. The missing values (lines) in the $k$-space domain are interpolated using information from the adjacent acquired lines in the $k$-space. The interpolation of missing $k$-space data (measurements), denoted by $\hat{s}_l(m,n)$, can be expressed through the following equation:

$$\hat{s}_l(m,n) = \sum_{l'=1}^{L} \sum_{m'=-M/2}^{M/2-1} \sum_{n'=\{-1,1\}} s_l(m-m', n-n')\omega_{l'l}(m',n') \tag{2.40}$$

with $M$ chosen such that $(M/2-1) = \lfloor k_{xmax}/\Delta k_x \rfloor$, where maximum attainable $k$-space point in the $x$ direction is denoted as $k_{xmax}$ and $\lfloor a \rfloor$ indicates the greatest integer that is less than or equal to $a$ [Ramos Llordén 18]. $\omega_{l'l}(\cdot,\cdot)$ is the GRAPPA kernel which defines

how the missing $\mathscr{k}$-space data in the $l$-th coil are interpolated using the acquired $\mathscr{k}$-space data from the $l'$-th coil. The weights for the GRAPPA kernel are derived from the so-called auto calibration signal (ACS) lines which cover the low-frequency spectrum of the $\mathscr{k}$-space data of every coil [Griswold 02]. Once the interpolation of $\mathscr{k}$-space data points is done for each coil measurements, the IFFT is employed for reconstruction of coil images $\hat{\mathbf{x}}_l$. The final composite image $\hat{\mathbf{x}}_{comp}$ can be estimated with the sum of squares (SoS) technique or with the spatially matched filter (SMF) [Roemer 90]. As a result with SoS we always obtain a real and positive $\hat{\mathbf{x}}_{comp}$ image. The SMF is known to be a computationally expensive technique and there are several variations of this technique where filter coefficients are estimated adaptively [Walsh 00].

### 2.2.6 LORAKS method

A recent approach in MR reconstruction considers low-rank modeling of local $\mathscr{k}$-space neighborhoods [Haldar 14,Haldar 16,Kim 17]. If the image support does not occupy the entire $FOV$, which is the usual case for MRI images, and/or if the image has slowly varying phase, then it is possible to arrange image Fourier coefficients ($\mathscr{k}$-space data) into a matrix in such way that it has low-rank [Haldar 14]. One way is to select the local neighborhood of $\mathscr{k}$-space data points from $\mathbf{y}$ within some distance from the neighborhood center, which is arbitrarily selected in the $\mathscr{k}$-space, and to place that neighborhood into a row of matrix denoted by $\mathbf{H}$. This ordering of $\mathscr{k}$-space data points results in a rank-deficient Hankel structured matrix. Frequently, the rank of $\mathbf{H}$ will become smaller as the support of the image decreases. There is also another way how the $\mathbf{H}$ can be created where the rank of the matrix can be associated with the image phase e.g. rank becomes smaller as the image phase gets smoother [Haldar 16]. As a result, encouraging lower rank of the Hankel-like matrix $\mathbf{H}$ during image reconstruction provides that recovered image has smaller spatial support or slow varying phase or both. A recently described constrained image reconstruction framework designed for single-channel image data named LORAKS [Haldar 14,Haldar 16,Kim 17] utilizes this fact and exploited it to reduce sampling requirements. LORAKS uses regularization terms in objective function in order to control how strongly the rank constraints are imposed. This will lead to a reconstructed image with limited spatial support and/or slowly varying phase. The regularization term is defined as:

$$R(\mathbf{y}) = \|\mathbf{H} - \mathbf{H}_r\|_F^2 \tag{2.41}$$

where $\|\cdot\|_F$ denotes the Frobenius norm and $\mathbf{H}_r$ is the optimal low-rank matrix approximation of $\mathbf{H}$. With the user-defined rank parameter $r$, the matrix $\mathbf{H}_r$ is obtained by truncating the singular value decomposition (SVD) of $\mathbf{H}$ at rank $r$. Although the cost function for LORAKS is non-convex due to the involved low-rank regularization term, local optima can be obtained by a simple majorize-minimize algorithm that is guaranteed to monotonically decrease the cost function value until it converges [Haldar 14]. In Chapter 6 we will compare our developed methods with extensions of LORAKS, P-LORAKS [Haldar 16] and SENSE-LORAKS [Kim 17] methods developed for reconstructing parallel imaging data while simultaneously leveraging support and phase constraints.

### 2.2.7 Conclusion

In this chapter, we introduced basic physical principles behind the creation of MR images. Following the principles, a mathematical formulation of the MR image reconstruction problem is given in the form of a linear inverse Fourier problem or as an optimization problem. We further reviewed traditional methods for MRI image recovery derived from the stated problem formulation when the single or multi-coil approach in MRI acquisition is used. In the following chapter, we will consider MR reconstruction from undersampled data using the compressed sensing (CS) theory. We will review relevant state-of-the-art methods from the field which are solving the unconstrained optimization problem stated in (2.32) or its constrained formulation with different involved regularization.

# 3

# MRI recovery from partial data

*Intelligence is not only the ability to reason. It is also the ability to find relevant material in memory and to deploy attention when needed.*
—Daniel Kahneman

This chapter reviews the principles and representative approaches for MR image recovery from undersampled measurements. These approaches are in the literature often referred to as compressed sensing (CS). We start from the general concept of sparse MRI reconstruction and the fundamental principles of the CS theory. We then describe sparsifying transforms which are commonly used for sparse signal representation, which are also used in our algorithms in the subsequent chapters. The chapter ends with a review of current state-of-the-art CS-MRI methods, some of which will serve as a basis for our techniques in the following chapters.

## 3.1 Sparse MRI

The speed of image acquisition in MRI is fundamentally limited by physical (gradient amplitude and slew-rate) and physiological (nerve stimulation) constraints. A sampling of $k$-space data below the Nyquist sampling rate leads to aliasing artifacts in the Fourier reconstruction (see Fig. 3.1). Therefore reducing the amount of $k$-space data, thus increasing the imaging speed, without degrading the image quality is the research subject considered by many researchers.

Many proposed approaches in the literature try to mitigate undersampling artifacts. They can be classified into three groups: i) methods generating artifacts that are incoherent or less visually apparent

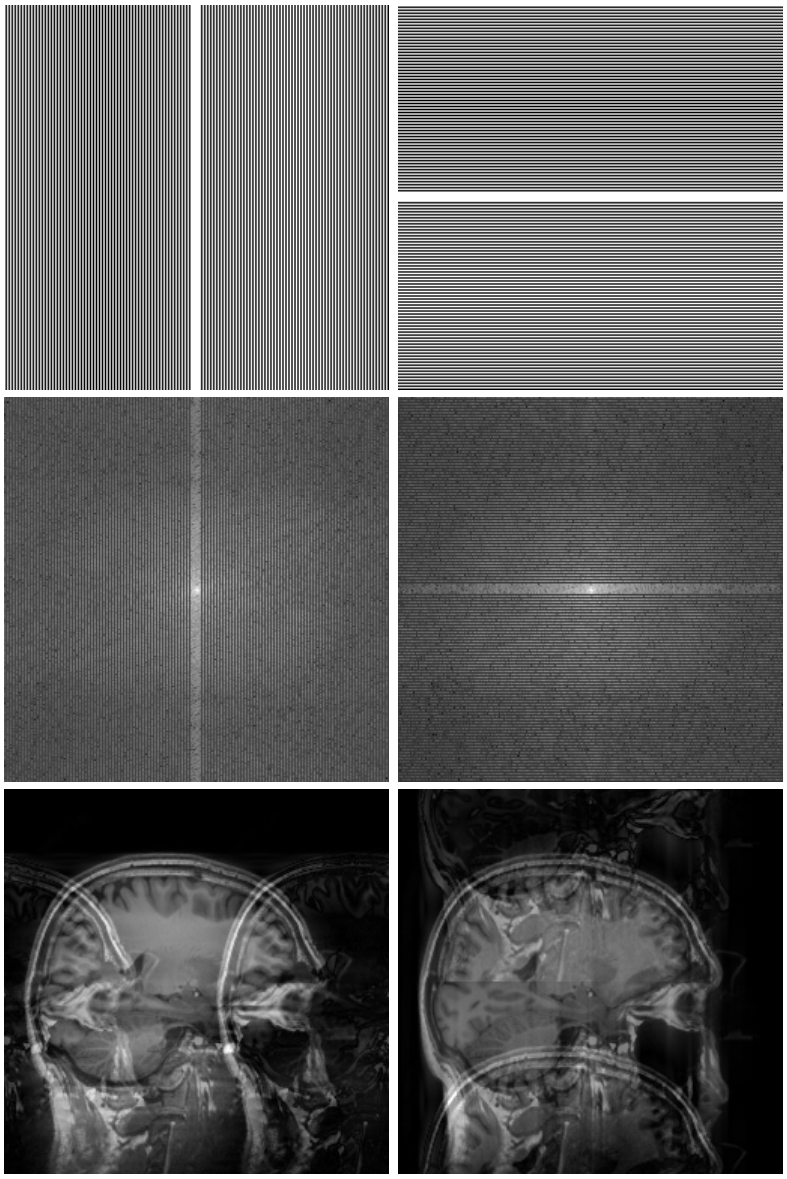


**Figure 3.1:** Aliasing artifacts caused by undersampling in Fourier domain in **Left**: $k_x$ and **Right**: $k_y$ direction. From top to bottom: undersampling pattern, the magnitude of undersampled Fourier coefficients, reconstructed image from undersampled coefficients (51%) using IFFT transform (49% of non-sampled coefficients are set to zero in order to use IFFT).

[Tsai 00,Marseille 96,Greiser 03], ii) methods exploiting redundancy in $k$-space such as parallel imaging [Pruessmann 99, Griswold 02, Blaimer 04] and iii) methods exploiting either spatial or temporal redundancy [Korosec 96, Tsao 03, Mistretta 06].

The sparsity, as a signal property can be exploited to significantly undersample the $k$-space during MR acquisition. Some types of MR images such as angiograms are already sparse in the spatial domain (pixel representation). This means that there are relatively few significant pixels with non zero values. However, most of the MR images tend to have a sparse representation in a well-chosen transformation domain such as finite differences or wavelet-like transform. This type of sparsity, in a known and fixed mathematical transform domain, is often called implicit sparsity or transform sparsity. In the work of Lustig at all [Lustig 07] authors explore implicit sparsity of MR images in the development of the MR reconstruction method from undersampled measurements. By involving implicit sparsity they consider spatial and temporal image redundancy in reconstruction procedure.

The motivation for including a sparsity constraint in the MR reconstruction, in particular exploiting the transform sparsity, comes from the image compression research field and widespread success of compression techniques. The most used and well-known are JPEG and JPEG-2000 which utilize discrete cosine transform and wavelet transform, respectively. These transformations represent the image content as a vector of coefficients, of which only relatively few are large enough and thus significant to store for later decoding and image reconstruction. Natural images, including medical images as well, are typically implicit sparse. Nevertheless, sparsity as a powerful constraint proved in image compression raised a very simple and intuitive question: do we need to acquire all those data in the first place if we plan to compress them afterward?

Development of the mathematical theory of compressed sensing (CS) in recent years opens up new possibilities for improving MRI reconstruction in terms of speed and image quality. Following the mathematical results in CS theory, if the image is sparse in a transform (spatial) domain and if $k$-space undersampling results in incoherent artifacts in that transform (spatial) domain, then the image can be recovered from undersampled frequency domain data with an appropriate non-linear reconstruction method. In the following sections, we give a short introduction to the CS theory and sparse signal representation providing the necessary basis for the understanding state-of-the-art in MRI reconstruc-

tion.

## 3.2 Principles of Compressed Sensing

The Compressed Sensing (CS) theory brings a new sampling principle that reshapes the digital signal processing, by revealing that it may be possible to surpass the traditional limits of the sampling theory established by pioneering works of Kotelnikov, Nyquist, Shannon, and Whittaker [Kotelnikov 33, Nyquist 28, Shannon 49, Whittaker 15]. These classical works established that signals can be exactly recovered from a set of uniformly spaced samples taken at the Nyquist rate of twice the highest frequency present in the signal of interest. On the wings of this discovery, much of signal processing has moved from the analog to the digital domain. Digitalized sensing and processing systems proved to be more robust, flexible and cheaper compared to their analog counterparts and led to a huge increase in the amount of data generated. In some important applications, the Nyquist rate can be so high that it can be simply physically impossible to build devices that acquire samples at the necessary rate. The resulting challenges in the acquisition and processing of signals are especially prominent in application areas such as imaging, video, medical imaging, remote surveillance, spectroscopy, and genomic data analysis.

Dealing with high-dimensional data or large amounts of data usually requires some kind of compression. The compression techniques provide a more concise representation of a signal possibly at the expense of some distortion (in the case of lossy compression). Transform coding is a popular compression approach that relies on finding a basis or frame that provides sparse or compressible representation for signals [Bruckstein 09, Protter 07]. The representation is called sparse if for a signal of length $N$ only $K \ll N$ coefficients are non-zero. In a compressible representation, the signal is well-approximated by $K \ll N$ non-zero coefficients.

Motivated by the concept of transform coding, CS introduced a new framework where data compression takes place already during signal acquisition. According to theory a large reduction in sampling and computation costs for sensing is possible for signals that admit a sparse or compressible representation in a well-chosen transform domain. The theory of CS was formalized in the seminal works of Candès, Romberg, Tao and Donoho [Candès 06a, Candes 06c, Candès 06d, Candes 06e, Candes 06f, Baraniuk 07, Donoho 06a]. They demonstrate that a band-limited

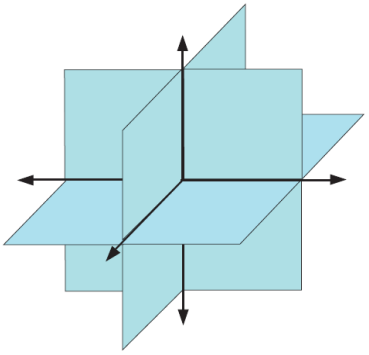


**Figure 3.2:** Union of subspaces defined by $\sum_2 \subset \mathbb{R}^N$, i.e, the set of all 2-sparse signals in $\mathbb{R}^3$.

signal having sparse or compressible representation can be fully recovered from a small set of linear, nonadaptive measurements provided that they are incoherently undersampled. This led to a fundamental idea behind CS: rather than sampling at a high rate and then compressing the sampled data, we have the possibility to directly sense the data in a compressed form, i.e. below the Nyquist sampling rate.

In mathematical terms, a high-dimensional data (image) can be vectorized in $\mathbf{x} \in \mathbb{R}^N$ and represented in a basis/dictionary $\mathbf{\Psi} \in N \times N$ as:

$$\mathbf{x} = \mathbf{\Psi}\boldsymbol{\theta} \tag{3.1}$$

where $\boldsymbol{\theta}$ are the coefficients of the new signal representation. If the signal representation is sparse in the chosen basis/dictionary $\mathbf{\Psi} \in \mathbb{R}^{N \times N}$, then only $K \ll N$ components of $\boldsymbol{\theta}$ are non-zero i.e. $\|\boldsymbol{\theta}\|_0 = K$. A geometrical explanation is that sparse signals live in the union of all possible $\binom{N}{K}$ subspaces denoted as $\sum_K$. In Fig. 3.2 we show $\sum_2$ embedded in $\mathbb{R}^3$, where we have $\binom{3}{2} = 3$ possible subspaces.

An important observation in practice is that very small number of real-world signals are truly sparse. They are rather compressible meaning that they can be well-approximated by sparse signals. For compressible signals the coefficients sorted in the order of decreasing magnitudes $|\theta_1| \geq |\theta_2| \geq |\theta_3| \geq ... \geq |\theta_N|$, obey a power law decay:

$$|\theta_i| \leq Ri^{-\frac{1}{r}}, \quad i = 1, 2...., N \tag{3.2}$$

where $R > 0$ and $r > 0$ are constants [Cevher 10a]. Due to this rapid

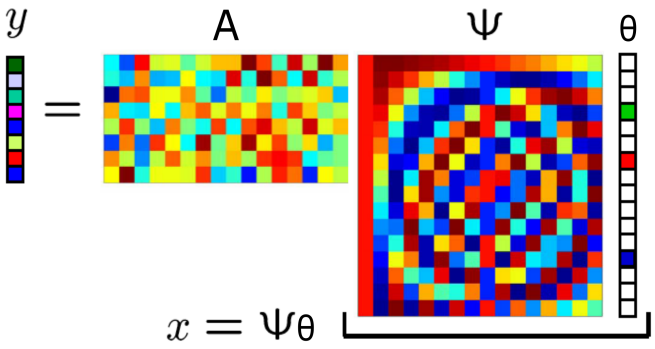



**Figure 3.3:** Compressive sensing measurements process with a sensing (measurement) matrix $\mathbf{A}$ and transform matrix $\mathbf{\Psi}$. The coefficient vector $\boldsymbol{\theta}$ is sparse with $K = 3$.

decay of the coefficients magnitudes, compressible signals can be represented accurately by $K \ll N$ coefficients.

In the following, we will restrict our attention to a standard finite dimensional CS model. In particular, a signal $\mathbf{x}$ or its representation $\boldsymbol{\theta}$ in a suitable basis/dictionary $\mathbf{\Psi}$ is not acquired directly by the sensing system. Instead, a measurement system creates $M$ linear projections $\mathbf{y} \in \mathbb{R}^M$, which is mathematically represented as:

$$\mathbf{y} = \mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{\Psi}\boldsymbol{\theta} \tag{3.3}$$

where $\mathbf{A}$ is $M \times N$ sensing (measurement) matrix. The matrix $\mathbf{A}$ is a dimensionality reduction operator since it projects from $\mathbb{R}^N$, where $N$ is generally large, to $\mathbb{R}^M$ where $M \ll N$. Fig. 3.3 illustrates the process of acquiring measurements. In the adopted CS framework, we assume that the projections (measurements) are non-adaptive meaning that the rows of $\mathbf{A}$ are fixed in advance and do not depend on the previously acquired measurements. The sensing matrix $\mathbf{A}$ needs to satisfy some conditions in order to sustain the unique information for some particular $K$-sparse or compressible signal after the linear projection (dimensionality reduction). This property of the matrix $\mathbf{A}$ together with the sparsity or compressibility of $\mathbf{x}$ is necessary for distinguishing $\mathbf{x}$ among all possible solutions of the undetermined system in (3.3). A restricted isometry property (RIP) is commonly used criterion for evaluating the desired properties of matrix $\mathbf{A}$ [Candès 06a]. It requires that $\mathbf{A}$ approximately preserves the distances between all signal pairs in the set (subspace $\sum_K$) of sparse signals. Formally, a $M \times N$ measurement matrix $\mathbf{A}$ has the

$K$-restricted isometry property (K-RIP) with constant $\epsilon_K < 1$ if for all $K$-sparse signals $\mathbf{x} \in \sum_K$,

$$(1 - \epsilon_K)\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1 + \epsilon_K)\|\mathbf{x}\|_2^2. \tag{3.4}$$

This means that the measurement matrix $\mathbf{A}$ accomplishes stable embeddings for an arbitrary sparse vector by not altering its magnitude (the $\ell_2$). If the matrix $\mathbf{A}$ satisfies the $2K$-RIP, then from (3.4) follows that matrix $\mathbf{A}$ approximately preserves the distance between any pair of $K$-sparse signals. This allows a unique $K$ sparse signal to be recovered in the absence of noise, for a given measurement matrix $\mathbf{A}$ which satisfies the $2K$-RIP [Cheng 16, Candes 08]. There are other conditions on $\mathbf{A}$ besides K-RIP which are required in some different contexts such as null space property, spark, and unique representation property [Cohen 09]. Random matrices will satisfy K-RIP with high probability if the entries are independent and identically distributed (i.i.d.) according to any sub-Gaussian distribution provided $M = \mathcal{O}(N \log(N/K))$ [Candès 06a, Donoho 06a]. They also satisfy a so-called universality property in that, for any choice of orthonormal basis matrix $\mathbf{\Psi}$, the matrix $\mathbf{A}\mathbf{\Psi}$ will also satisfy the K-RIP property with high probability. A natural generalization of the K-RIP property, considering highly overcomplete (redundant) dictionaries $\Psi \in \mathbb{R}^{N \times D}$ with $N \ll D$ was introduced in [Candes 11]. This theoretical result allows the utilization of tight frames (redundant dictionaries) in the CS framework and also offers convenience and flexibility induced by overcomplete representations of the signal. The flexibility comes from the fact that there are many ways to reconstruct the original image from its overcomplete representation. Further, in practice, there are numerous examples in which a signal of interest is not sparse in an orthonormal basis but in an overcomplete dictionary [Candes 11]. A translation-invariance property is a convenience that comes with overcomplete representation, which is lost in orthonormal representation. This property turns out to be extremely helpful in applications such as filtering, deconvolution, detection, or more generally, analysis of data [Starck 04, Starck 07].

In the next subsection, we give an overview of formal approaches for signal reconstruction based on the optimization theory.

### 3.2.1 Tractable recovery algorithms

Recovering the signal $\mathbf{x} = \mathbf{\Psi}\boldsymbol{\theta}$ from the acquired (partial) measurements $\mathbf{y}$ demands involving some prior knowledge about the signal (either spar-

sity of compressibility) since there are a lot of possible solutions for undetermined system in (3.3). When we recover a strictly sparse signal from the measurements without the presence of noise, we are looking for the sparsest $\boldsymbol{\theta}$ that agrees with the measurements $\mathbf{y}$. With the $\ell_0$ pseudonorm that counts the number of non-zero entries in $\boldsymbol{\theta}$, the optimization problem for sparse signal reconstruction is:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \|\boldsymbol{\theta}\|_0 \quad \text{s.t.} \quad \mathbf{A}\boldsymbol{\Psi}\boldsymbol{\theta} = \mathbf{y}. \tag{3.5}$$

Although the optimization in (3.5) can recover a $K$-sparse signal from $M = 2K$ compressive measurements [Baraniuk 10, Cevher 10b], its solving is unfortunately a combinatorial $NP$-hard problem and also not stable in the presence of noise. One direction for translating this problem into a more tractable one is to replace the $\ell_0$ pseudo-norm with the $\ell_1$ norm, which yields a convex optimization problem:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \|\boldsymbol{\theta}\|_1 \quad \text{s.t.} \quad \mathbf{A}\boldsymbol{\Psi}\boldsymbol{\theta} = \mathbf{y}. \tag{3.6}$$

The optimization problem in (3.6) is known as basis pursuit (BP) and corresponds to a linear program that can be solved in a polynomial time [Candès 06a,Donoho 06a,Baraniuk 10]. In [Candes 08] author states that the $\ell_0$ and $\ell_1$ problem formulations from (3.5) and (3.6) respectively, are formally equivalent under the $2K$-RIP condition given in (3.4), with the frame bound constant $\epsilon_{2K}$. The difference is in the frame bound condition constant for the perfect reconstruction of a $K$-sparse signal, which is $\epsilon_{2K} < 1$ for the $\ell_0$ problem formulation, and $\epsilon_{2K} < \sqrt{2} - 1$ for the $\ell_1$ problem formulation. In situation where we have to deal with additive noise in (3.3), different adaptations of the problem formulations in (3.6) are proposed, leading to methods such as basis pursuit with denoising (BPDN) [Chen 98] and the least absolute shrinkage and selection operator (LASSO) [Figueiredo 07]. As an alternative to convex optimization based algorithms, greedy methods are used which directly solve the $\ell_0$ problem formulation from (3.5). They are conducting iterative greedy search for the sparsest solution $\boldsymbol{\theta}$, which agree with the measurements $\mathbf{y}$, by thresholding the signal $\boldsymbol{\theta}$, that is, by setting to zero all except the $K$ largest entries of $\boldsymbol{\theta}$. The most prominent greedy based algorithms are: iterative hard thresholding (IHT) [Blumensath 09], compressive sampling matching pursuit (CoSaMP) [Needell 09a] and subspace pursuit (SP) [Dai 09]. Interestingly, the algorithms from both groups have similar theoretical recovery guarantees from $M = \mathcal{O}(N \log(N/K))$ measurements, when the measurement matrix satisfies the $2K$-RIP property with an algorithm dependent $\epsilon_{2K}$.

Extension of iterative thresholding algorithms is proposed under an approximate message passing (AMP) framework [Maleki 10]. AMP is employed for solving the adaptation of the problem in (3.6) when the presence of additive Gaussian noise in measurements is considered. Under the AMP framework authors introduced the Onsager correction term, derived from the theory of belief propagation in graphical models, which at each iteration approximately Gaussianizes the residual error [Metzler 16]. Recently an improved, more robust versions of the AMP algorithm are proposed: vector AMP (VAMP) [Rangan 19], which has considered the case of ill-conditioned measurement matrices, and a generalized AMP (GAMP) [Rangana 19, Vila 15] where damping and mean-removal procedures have been utilized to ensure algorithm convergence.

In the following section, we introduce sparsifying transforms for images that we shall employ in our work. We start with the orthogonal discrete wavelet transform. We then review its non-decimated version and explain improvements which it brings in signal representation. To improve directionality in sparse signal representation compared to wavelet transform we introduce a non-decimated shearlet transform which will be used later on in our algorithm development.

## 3.3 Sparse signal representation

Signal representation plays a crucial role in the extraction of useful information from various signal modalities (audio, image, video). By transforming a signal into a more suitable domain for a particular task, the extraction of signal features of interest can be largely facilitated. The Fourier transform decomposes signals using harmonic functions, which are the basis functions of this transform. Hence the Fourier transform reveals the spectral content of the signal. For uniformly regular signals the Fourier coefficients are sparse [Mallat 99].

If we restrict ourselves to discrete 2D signals (images), here denoted as $f(m,n)$ where $(m,n)$ represents the coordinates of the pixel position, the Fourier transformation is defined as

$$F(u,v) = \sum_{m=1}^{M}\sum_{n=1}^{N} f(m,n)e^{\left[-2\pi i(\frac{mu}{M}+\frac{nv}{N})\right]}, \qquad u = 1,2,3,...,M. \quad v = 1,2,3,...,N. \tag{3.7}$$

while the inverse Fourier transform is given by

$$f(m,n) = \frac{1}{MN}\sum_{u=1}^{M}\sum_{v=1}^{N} F(u,v)e^{\left[2\pi i(\frac{mu}{M}+\frac{nv}{N})\right]}, \qquad m = 1,2,3,...,M. \quad n = 1,2,3,...,N. \tag{3.8}$$

Variables $(u,v)$ are called spatial frequencies in analogy with 1D Fourier transform terminology [Woods 11, Sonka 14]. In the equation (3.8) the discrete image $f$ is represented as a linear combination of simple periodic patterns $e^{2\pi i(\frac{mu}{M}+\frac{nv}{N})}$ which form an orthogonal basis. The coefficients $F(u,v)$, used as weights in (3.8), are obtained as a linear combination of projections of image $f$, defined in (3.7), onto the basis vectors.

The Fourier representation is used in signal processing as an important tool for analyzing linear, time-invariant systems and for a wide range of applications such as signal transmissions or stationary signal processing [Mallat 99]. However, when it comes to representing transient phenomena (localized events in a signal), the Fourier transform becomes inadequate. It provides only the global information about the spatial frequencies for the whole image but no information about where in the image they are present. Hence, Fourier coefficients $F(u,v)$ lose all spatial interpretation and represent a particular frequency contained in the whole spatial domain of the image. As such global transform, it is not adequate for the analysis of non-stationary signals containing spikes or edges at arbitrary positions.

The Gabor-Heisenberg uncertainty principle determines the relation between frequency and time resolution of the 1D signal. According to the uncertainty principle, a signal cannot be perfectly localized both in frequency and in time domain simultaneously [Amrein 77, Gabor 46]. This principle is also valid for the signal originally defined in the spatial domain (images), i.e. for their spatial-frequency decomposition. Therefore any attempt to make a frequency space localizing transform implies a loss of spatial compactness, so we can only optimize the trade-off.

One way to localize changes in time for a 1D signal of length $M$ is to use short time Fourier transform (STFT):

$$F(t,u) = \sum_{m=1}^{M} f(m)w(m-t)e^{-2\pi i\frac{mu}{M}} \tag{3.9}$$

the extension of which for multidimensional signals given in [Hinman 84]

can be used for images:

$$F(g,h,u,v) = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} f(m,n)w(g-m,h-n)e^{\left[-2\pi i(\frac{mu}{M}+\frac{nv}{N})\right]}, \quad u = 1,2,3,...,M. \quad v = 1,2,3,...,N. \tag{3.10}$$

Here the signal is divided into small short-time (space) sections using the window function $w(\cdot)$ and treated locally as it were periodic. This results in two (four) dimensional representation for 1D (2D) signals through $F(t,u)$ and $F(g,h,u,v)$ coefficients, respectively. The choice of the support for the window function influences the signal representation making a compromise between time (spatial) and frequency resolution. A narrow support for the window function $w(\cdot)$ yields lower frequency resolution but better localization in time (space), while a wide support leads to a worse localization in time (space) but better frequency resolution. The wavelet transform largely overcomes this shortcoming of the STFT transform by employing different basis functions with variable support size which offers multiresolution analysis of the signal and therefore offers better localization more faithful representation.

In image processing, transients (e.g. image discontinuities) contain the most essential information in the signal and it is of great importance to have a representation that is able to localize well both the spatial frequencies and the spatial information in a signal. The wavelet and wavelet-related transforms with predefined basis functions raised as very efficient and prominent in image representation [Jacques 11]. In the following subsections, we give an introduction to 2D discrete wavelet transform (DWT) and we review its advantages over the 2D discrete Fourier transform. Then we review the non-decimated version of DWT, which will be utilized in our MRI signal reconstruction algorithms as a sparsifying transform.

### 3.3.1 Wavelet analysis

The wavelet transform is a tool to analyze signal according to scale by expanding it into different frequency components. A detail mathematical foundation of the wavelet transform can be found in [Daubechies 88, Mallat 89b,Daubechies 92,Chui 16,Mallat 99,Cohen 96]. What makes the wavelet transform unique and creates a difference in comparison to the Fourier transform or its short-time version (STFT), is better time (space)

frequency localization and representation of local transient structures in a signal or its singularities. For example, in images, structures like edges or irregular textures are represented with a few large wavelet coefficients located in their neighborhood.

A family of wavelet basis functions, i.e., 'wavelets' are generated from a unique prototype function so-called 'mother wavelet' [Daubechies 92, Mallat 99] by shifts and dilatations. They are called wavelets since they resemble small waves, providing localization at multiple signal scales in time (space) to a certain degree due to the uncertainty principle [Mallat 99]. A wavelet basis for the $\mathbf{L}^2(\mathbb{R})$ Hilbert space of finite energy functions $f$ is constructed as

$$\left\{\psi_{j,k}(x) = \frac{1}{\sqrt{2^j}}\psi\left(\frac{x - 2^j k}{2^j}\right)\right\}_{j\in\mathbb{Z},k\in\mathbb{Z}}, \tag{3.11}$$

where $x \in \mathbb{R}$ and $\psi_{j,k}(x)$ are suitably constructed from the mother wavelet $\psi(x)$ using scaling (dilatation) defined with $2^j$ and translation realized with $2^j k$, $k \in \mathbb{Z}$. The support of wavelets $\psi_{j,k}(x)$ is proportional to the scale $2^j$. The mother wavelet function $\psi(x)$ needs to satisfy the admissibility condition:

$$C_\psi = \int_{\mathbb{R}} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty \tag{3.12}$$

where $\Psi(\omega) = \int_{\mathbb{R}} \psi(x) e^{-ix\omega} dx$ is the Fourier transform of $\psi(x)$. Admissibility condition implies that $\Psi(0) = \int_{\mathbb{R}^2} \psi(x) dx = 0$. This means that $\psi(x)$ must be oscillatory and well localized (i.e., rapidly decreases to zero when $|x|$ tends to infinity in both dimensions in the case of two dimensional wavelets). By convention, it is centered around $x = 0$, and has a unit norm ($\|\psi(x)\| = 1$).

The origins of the wavelet analysis can be traced back to the work of Alfred Haar [Haar 10] who studied orthogonal function spaces and to the work of Grossmann and Morlet [Grossmann 84] who set the basis of signal analysis using wavelets. Among the various constructions of an orthogonal wavelet basis, including [Haar 10, Meyer 89, Daubechies 88, Morlet 82], the Daubechies wavelets [Daubechies 88] have the best properties (shortest wavelets for a given number of vanishing moments). The orthogonality property puts a strong limitation on the construction of wavelets. The Haar wavelet is the only real-valued wavelet that is compactly supported, symmetric and orthogonal [Daubechies 88]. Therefore, a generalization to biorthogonal wavelets has been considered to gain more flexibility in design a wavelet basis functions.

### 3.3.2 Discrete wavelet transform

The discrete wavelet transform (DWT) decomposes a signal in a (bi-) orthogonal wavelet basis. Any finite energy signal $f(x)$ can be decomposed as:

$$f(x) = \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} w_{j,k}\psi_{j,k}(x) \tag{3.13}$$

where $w_{j,k}$ are the wavelet coefficients given by the inner product of $f(x)$ with the dual basis functions $\widetilde{\psi}_{j,k}$:

$$w_{j,k} = \langle f, \widetilde{\psi}_{j,k}\rangle = \sum_{m=1}^{M}\sum_{n=1}^{N} f(m,n)\widetilde{\psi}_{j,k}(m,n) \tag{3.14}$$

where $\widetilde{\psi}_{j,k}(m,n) = 0$ outside its support. If dual basis functions $\widetilde{\psi}_g$, are orthogonal to the corresponding functions $\psi_h$ of the expansion set, where $g = (j,k)$ and $h = (i,z)$, then $\widetilde{\psi}_g$ is biorthogonal to $\psi_h$. This means that $\langle \psi_h, \widetilde{\psi}_g\rangle = \delta(h-g)$ where $\delta(i)$ is the Kronecker delta. The wavelet coefficient $\langle f, \widetilde{\psi}_{j,k}\rangle$ has a small amplitude $|\langle f, \widetilde{\psi}_{j,k}\rangle|$ if $f$ is regular over the support of $\widetilde{\psi}_{j,k}$. The large amplitudes are present near sharp transitions such as edges.

The orthogonal wavelets are rarely available as closed form expressions, but rather obtained through a computational procedure which uses discrete filters. Stéphane Mallat through the elaboration of multiresolutional signal approximation in [Mallat 89a, Mallat 89b] proposed a systematic framework for constructing wavelet bases. The multiresolutional theory proves that filter banks (a set of filters that decompose a signal into a set of components) can characterize a wavelet $\psi$ that forms an orthonormal basis for the $\mathbf{L}^2(\mathbb{R})$. This allows that a fast discrete wavelet transform be implemented by cascading filters from the filter banks [Mallat 89b]. Multiresolution signal analysis creates a series of approximations of a signal, at dyadic resolution scales. A family of scaling functions, defined as:

$$\left\{\phi_{j,k}(x) = \frac{1}{\sqrt{2^j}}\phi\left(\frac{x-2^j k}{2^j}\right)\right\}_{j\in\mathbb{Z},k\in\mathbb{Z}}, \tag{3.15}$$

are used to generate signal approximations at different scales. Signal approximation $f_j(x)$ at the scale $2^j$ is defined as:

$$f_j(x) = \sum_{k\in\mathbb{Z}} c_{j,k}\phi_{j,k}(x) \tag{3.16}$$

where $c_{j,k}$ are scaling coefficients obtained as the inner product of $f(x)$ with the dual scaling functions $\widetilde{\phi}_{j,k}$, i.e, $c_{j,k} = \langle f, \widetilde{\phi}_{j,k} \rangle$. Construction of wavelet functions starts from scaling functions which span the spaces of signal approximations. The wavelet functions encode the difference in information between adjacent approximations. For image decomposition in a wavelet bases a two dimensional scaling function $\phi(m,n)$ and three two-dimensional wavelets $\psi^{HL}(m,n)$, $\psi^{LH}(m,n)$ and $\psi^{HH}(m,n)$ are required. Each is the product of two one dimensional functions

$$\begin{aligned} \phi^{LL}(m,n) &= \phi(m)\phi(n) \\ \psi^{HL}(m,n) &= \psi(m)\phi(n) \\ \psi^{LH}(m,n) &= \phi(m)\psi(n) \\ \psi^{HH}(m,n) &= \psi(m)\psi(n) \end{aligned} \tag{3.17}$$

producing separable scaling and separable 'directionally sensitive' wavelets. The superscripts indicate the use of a high-pass wavelet filter (symbol $H$) or a low-pass scaling filter (symbol $L$) for the first dimension (the first subscript) or the second dimension (the second subscript). Applying the low-pass filtering in both dimensions leads to a scaling filter $\phi^{LL}$. These wavelets measure intensity variations in images along different directions: $\psi^{HL}$ measures variations in direction of columns (for example, horizontal edges), $\psi^{LH}$ is sensitive to variations in rows direction (like vertical edges), and $\psi^{HH}$ corresponds to joint variations in both directions. The two dimensional scaling and wavelet functions are defined as:

$$\begin{aligned} \phi^{LL}_{j,u,v}(m,n) &= \frac{1}{\sqrt{2^j}} \phi^{LL}(\frac{m-2^j u}{2^j}, \frac{n-2^j v}{2^j}) \\ \psi^{i}_{j,u,v}(m,n) &= \frac{1}{\sqrt{2^j}} \psi^{i}(\frac{m-2^j u}{2^j}, \frac{n-2^j v}{2^j}), \quad i \in \{HL, LH, HH\} \end{aligned} \tag{3.18}$$

where the index $i$ identifies the directional wavelets. The image $f(m,n)$ can be written as a series expansion of wavelet functions

$$\begin{aligned} f(m,n) = &\sum_u \sum_v c_{\phi^{LL}}(j_0,u,v) \phi^{LL}_{j_0,u,v}(m,n) \\ &+ \sum_{j=j_0}^{\infty} \sum_u \sum_v w_\psi(j,u,v) \psi_{j,u,v}(m,n) \end{aligned} \tag{3.19}$$

where superscript $i$ for filter directionality is omitted for simpler notation. For an arbitrary starting scale $j_0$, the coefficients $c_{\phi^{LL}}(j_0,u,v)$ define an

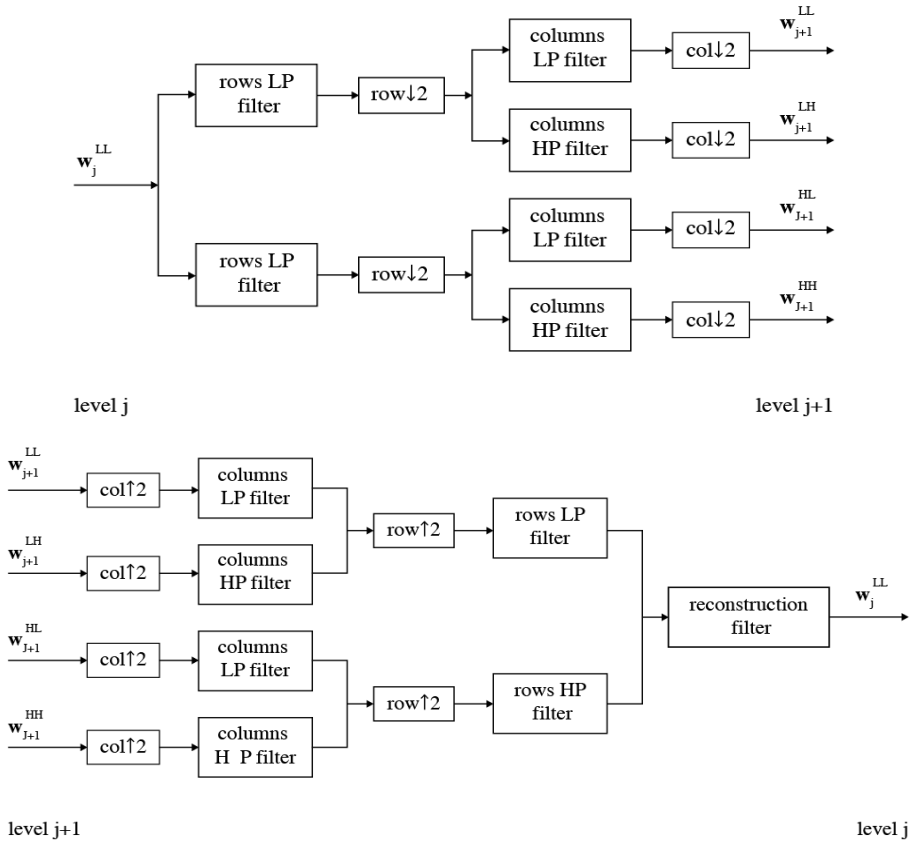



**Figure 3.4: Top**: 2D dicrete wavelet transform. A decomposition step at scale $j$ from the signal approximation $\mathbf{w}_j^{LL}$ **Bottom**: 2D inverse discrete wavelet transform. A reconstruction step at scale $j+1$ from wavelet coefficients $\mathbf{w}_{j+1}^{LH}, \mathbf{w}_{j+1}^{HL}, \mathbf{w}_{j+1}^{HH}$ and signal approximation $\mathbf{w}_{j+1}^{LL}$.

approximation of $f(m,n)$ at that scale. The $w_\psi(j,u,v)$ coefficients add horizontal, vertical, and diagonal details for scales $j \geq j_0$.

Fig. 3.4 shows one stage of a pyramidal filter bank procedure for computing all detail wavelet and approximation coefficients using orthogonal wavelets at level $j+1$ and reconstructing the signal approximation at level $j$. In the decomposition step, the input is convoluted with a low-pass (LP) filter for approximation coefficients and high-pass (HP) filter for detail wavelet coefficients. The outputs of both filtrations are then decimated by a factor of 2 in a direction opposite to the one in which filtration is conducted. After filtration in both directions followed by down-sampling (keeping only evenly indexed row or columns), an ap-

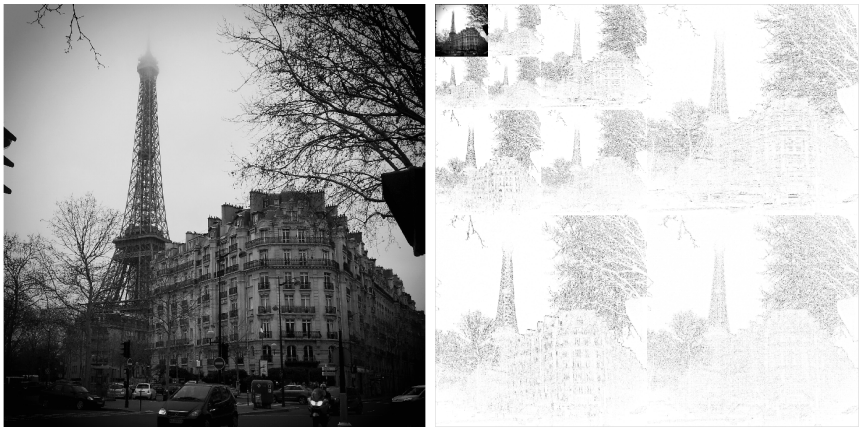


**Figure 3.5: Left** An image and its **Right** 2D discrete wavelet transform with Daubechies (db2) wavelet function. Black pixels denote large magnitude wavelet coefficients.

proximation $\mathbf{w}^{LL}$ and detail images $\mathbf{w}^{LH}, \mathbf{w}^{HL}, \mathbf{w}^{HH}$ at the particular scale are created. The decomposition at the next scale is obtained by repeating the process where the input is the approximation coefficients at the current scale. The reconstruction stage also involves low-pass and high-pass filtering but now with up-sampling (inserting zeros at odd-indexed rows or columns) with the same factor of 2. For a $N$-sample vector, the algorithm complexity is $\mathcal{O}(N)$, which is faster than FFT which needs $\mathcal{O}(N \log N)$ operations. In Fig. 3.5 we see a wavelet decomposition of an image in 4 scales and 3 orientations per scale. Each combination of scale and orientation corresponds to a particular subband. The amplitudes of the coefficients are shown, where the darker color corresponds to larger wavelet coefficients.

Wavelets have been widely used with enormous success in signal compression [Skodras 01,Antonini 92,Shapiro 93], feature extraction [Pittner 99,Lin 00,Bruce 02,Yen 00] and in noise suppression [Protter 07,Portilla 03, Dabov 07, Donoho 95, Pizurica 06]. The suppression of wavelet coefficients with small magnitudes can efficiently reduce the amount of noise while preserving the important image details. However, the classical (bi-)orthogonal discrete wavelet transform is not shift-invariant which makes it less suitable for pattern recognition [Mallat 96] and limits also performance in denoising applications [Coifman 95].

In the next subsection, we introduce a non-decimated (redundant) discrete wavelet transform which has a shift-invariance property and also due to the redundancy enables better estimation of the signal

from noisy observations.

### 3.3.3 Non-decimated discrete wavelet transform

The signal representation under the non-decimated discrete wavelet transform (NDWT) has the same number of wavelet coefficients at each scale. Formally we again have signal decomposition as in (3.13) but now the wavelets $\psi_{j,k}(x)$ are not linearly independent, i.e., they do not constitute a basis but a frame. If it is a tight frame then dual-frame is equal to the original wavelet frame $\psi(x) = \widetilde{\psi}(x)$. The NDWT can be found in literature under different names such as stationary wavelet transform [Nason 95], shift/translation invariant wavelet transform [Guo 95], redundant discrete wavelet transform [Fowler 05] or undecimated wavelet transform [Lang 96]. The main characteristic of NDWT is that it is redundant, shift-invariant, linear, and it gives a better approximation to the continuous wavelet transform than the approximation provided by the orthonormal discrete wavelet transform [Wang 10]. A nice overview can be found in [Fowler 05, Shensa 92].

The non-decimated wavelet transform is commonly implemented with the algorithme à trous [Holschneider 90, Mallat 92]. Instead of signal downsampling, this algorithm upsamples the filter responses by inserting zeros, i.e., 'holes' (in French trous) between the filter coefficients. This is done in both directions in case of a 2D transform. This transform requires more calculations and calls for bigger memory than orthogonal discrete wavelet transform and has computational complexity $\mathcal{O}(N \log N)$ [Mallat 91].

In Fig. 3.6, we illustrate the NDWT of the image shown in Fig. 3.5. The redundant wavelet frame decomposition enables better denoising quality which will be used in this thesis. In the next subsection, we introduce the non-decimated shearlet transform for images which improves a redundant signal representation by involving more directionalities across the scales.

### 3.3.4 Discrete shearlet transform

Wavelets are very efficient in approximating signals containing pointwise singularities. However for multidimensional data, different types of singularities are usually present and even more dominant. In images, sharp transitions like edges, contours and lines have various shapes and directionality. In order to express intrinsic geometrical features of multidimensional data one has to increase directional sensitivity of

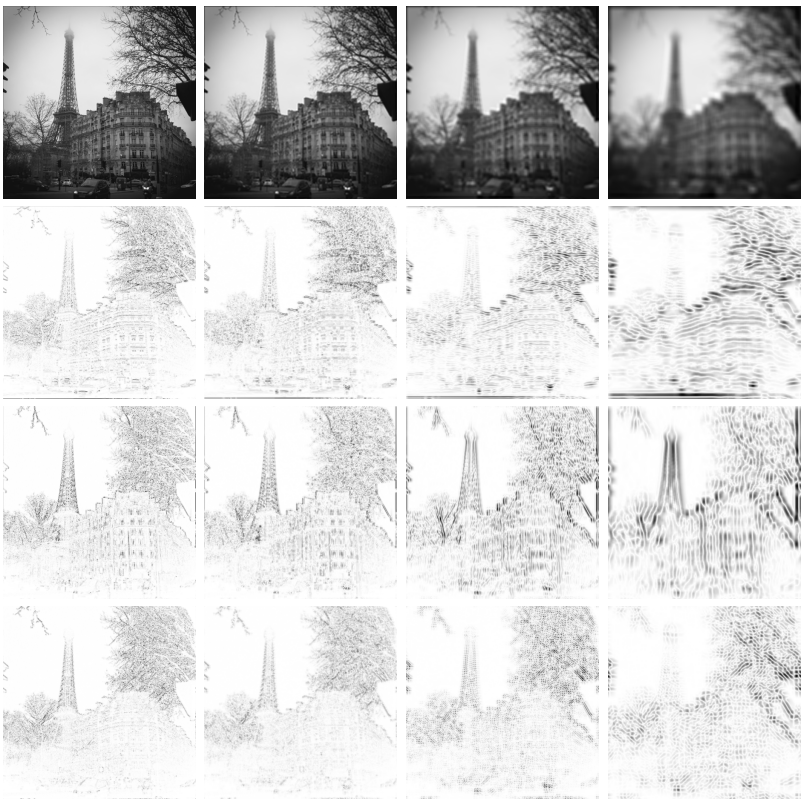


**Figure 3.6: Left to right** Different scales (from $1^{st}$ to $4^{th}$) of non-decimated discrete wavelet transform. **Up to bottom** Approximation coefficients followed by detail coefficients (horizontal, vertical and diagonal direction respectively). Black pixels denote large magnitude wavelet coefficients.

traditional wavelets. A variety of transformations which address this task have been proposed in literature such as: contourlet [Do 05], complex wavelets [Kingsbury 01], brushlets [Meyer 97], ridgelets [Candès 99], curvelets [Candès 04], bandelets [Le Pennec 05], and shearlets [Guo 06, Guo 07, Easley 08]. From the approximation theory point of view, all the mentioned transformations are searching for the representation system, a basis or more generally a tight frame, that achieves the best asymptotic decay rate of approximation error in signal representation. It has been shown that among design transformations only curvelets and shearlets are optimally sparse in representing 2-dimensional piecewise smooth functions $f(x)$ with discontinuities along twice continuously differentiable curves, i.e., $\mathbf{C}^2$ [Guo 07]. From this approximation

error associated with $N$-term reconstruction $f_N^S$ obtained by taking the $N$ largest coefficients in the shearlet expansion satisfies:

$$\|f - f_N^S\|_2^2 \leq BN^{-2}(\log N)^3, \quad N \to \infty \tag{3.20}$$

where $B$ is a constant. This property is often referred to as optimal sparsity [Guo 07].

The shearlet representation provides a flexible theoretical tool for the geometric representation of multidimensional data, which can be associated with multiresolution analysis [Labate 05] and is more natural for its discrete implementation compared to curvelets [Candès 04]. The shearlet transform also allows one to develop a variety of alternative implementations, with complete control over the mathematical properties of the transform [Easley 08, Goossens 09]. In the implementation of the shearlet transform, a so-called 'shearing' filters can have smaller support sizes than the directional filters used in the contourlet transform and can be implemented much more efficiently [Easley 08]. In the following, we will focus on a non-decimated version of the shearlet transform, which will be used as the sparsifying transform in some of our algorithms.

Shearlet functions $\psi_{j,l,k}$, where $j, l \in \mathbb{Z}, k \in \mathbb{Z}^2$ denote scale, orientation and spatial position respectively, are formed by dilating, shearing and translating the shearlet function $\psi \in \mathbf{L}^2(\mathbb{R}^2)$ as follows:

$$\psi_{j,l,k}(x) = |\det \mathbf{A}|^{j/2} \psi(\mathbf{B}^l \mathbf{A}^j x - k) \tag{3.21}$$

where $\mathbf{A}, \mathbf{B}$ are invertible $2 \times 2$ matrices with $|\det \mathbf{B}| = 1$. In this approach, the dilation matrices $\mathbf{A}^j$ are associated with scale transformations, while the matrices $\mathbf{B}^l$ are associated to area-preserving geometrical transformations, such as rotations and shear. In this thesis the normalization factor $|\det \mathbf{A}|^{j/2}$ has been chosen such that the norm $\|\psi\|_2 = \|\psi_{j,l,k}\|_2$ for all $j, l, k$ and we used the following transform matrices:

$$\mathbf{A} = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}. \tag{3.22}$$

The anisotropic scaling matrix $\mathbf{A}$ yields scaling operations in both directions performing a change of resolution, in particular, scaling in the $x$-direction is twice the scaling in the $y$-direction. The geometric shear matrix $\mathbf{B}$ conducts a linear map that transforms point coordinates $(x, y)$ into $(x + my, y)$, where $m = 1$ is the fixed-parameter called shear factor and represents a cotangent of the angle by which the vertical line tilts. Shearlets exhibit highly directional sensitivity compared to

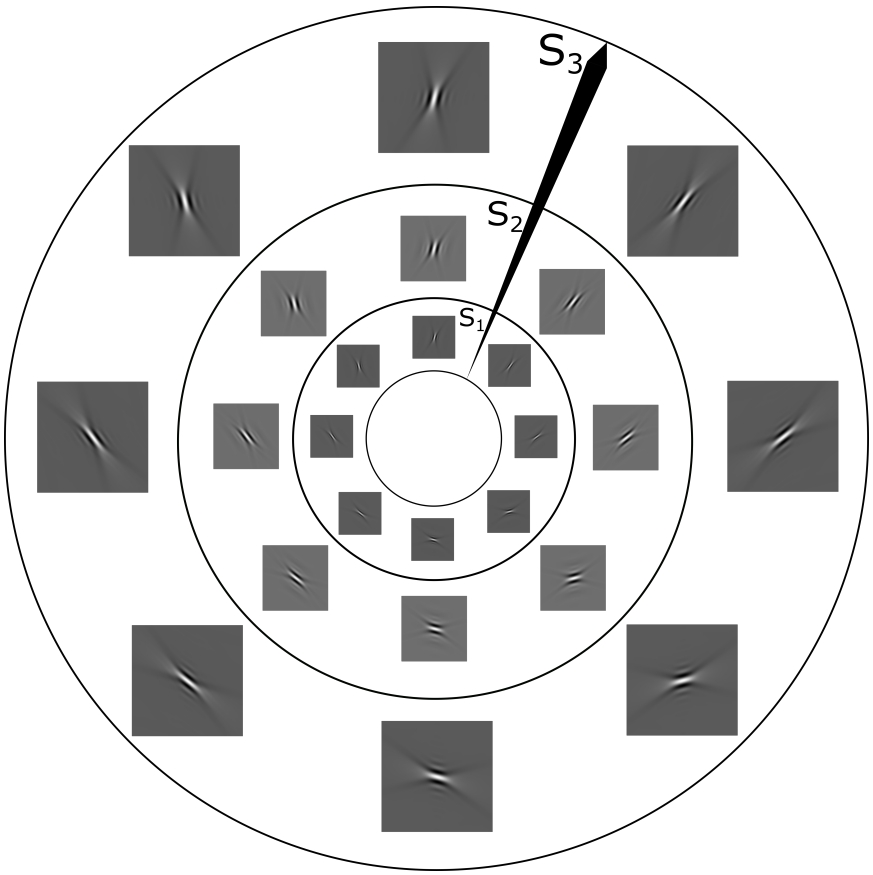



**Figure 3.7:** An illustration of the shearlet functions in a non-decimated 2D shearlet transform with 3 scales and 8 orientation per scale. Starting from the center towards the periphery in the direction of the arrow, between consecutive black circles are placed shearlet detail functions for the first, second and the third scale respectively (in image denoted with $S_1, S_2$ and $S_3$).

DWT with the possibility to have more than 3 orientations per each scale level. Fig. 3.7 illustrates shearlet functions used in a non-decimated 2D discrete shearlet transform. Corresponding shearlet decomposition, i.e., projections for the region of the image in Fig. 3.5, given in Fig. 3.7 are represented in Fig. 3.8.

Shearlet functions allow for a much less redundant tight frame signal representation compared to curvelets, while having shift-invariance property [Goossens 09]. The shearlet transform has been widely used in image denoising [Gibert 14, Ghofrani 15, Easley 08, Easley 09]. Clas-

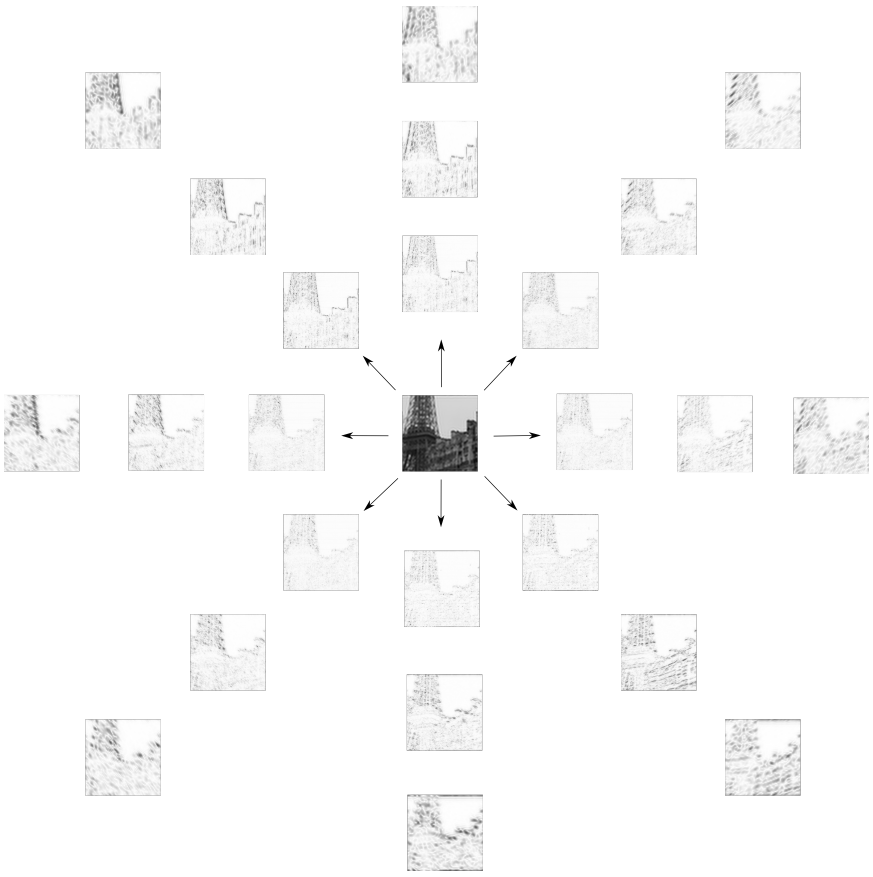


**Figure 3.8:** Shearlet decomposition with functions showed in Fig. 3.7. The position of each subband in the figure corresponds to the position of the shearlet functions in Fig. 3.7 used for calculation of the observed subband.

sical denoising methods with wavelet-like transforms are based on the application of thresholding rules on signal coefficients in the chosen basis or frame expansion. In [Lang 96] it was proved that thresholding rules applied on the coefficients obtained with non-decimated wavelet transform achieve better performances in noise reduction compared to the same thresholding rules applied on the decimated version of the wavelet transform. Other wavelet-like transform such as curvelet [Starck 02, Tessens 08] confirm also improvements in image denoising compared to the wavelet transform. In this thesis we consider non-decimated wavelet and shearlet transforms for signal representation. The decimated version of a wavelet-like transform creates a decorre-

lated set of coefficients for which i.i.d. (independent and identically distributed) models are justified and thus a derivation of a maximum a posteriori (MAP) or minimum squared error (MMSE) estimators of the signal are facilitated. In non-decimated wavelet-like transforms correlations between coefficients are present, which makes their statistical modeling far more difficult. However, non-decimated wavelet-like transforms enable simple interscale comparisons between the coefficients yielding better detection of useful image features.

In the following, we give an overview of the current state-of-the-art methods for solving the MRI reconstruction problem which involve CS theory and sparse signal representation (wavelet or wavelet-like transforms) and we also discuss possible improvements which will be incorporated and presented later in the following chapters.

## 3.4 CS based reconstruction algorithms

Slow data acquisition process in MRI [Wright 97, Lustig 07, Lustig 08] calls for development of smart undersampling schemes [Roman 14, Bastounis 15] and the corresponding reconstruction algorithms. Compressed Sensing (CS) [Candès 06b, Donoho 06b], demonstrated potential to improve the acquisition speed in MRI and since the seminal work of Lustig and collaborators [Lustig 07, Lustig 08] on CS-MRI, a number of studies including [Starck 05, Ma 08, Blumensath 09, Yang 10, Aelterman 11, Chen 12, Huang 11b, Rajani 12, Adcock 13, Chen 14], have addressed MRI recovery from partial data.

In a CS-MRI setup, the acquired $k$-space measurements $\mathbf{y} \in \mathbb{C}^M$ of an ideal image $\mathbf{x} \in \mathbb{C}^N$ are

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} \tag{3.23}$$

where $M \ll N$, $\mathbf{n} \in \mathbb{C}^M$ is white Gaussian noise, and $\mathbf{A} \in \mathbb{C}^{M \times N}$ denotes the undersampled Fourier operator [Lustig 07, Lustig 08]. Estimation of $\mathbf{x}$ from measurements $\mathbf{y}$ is an ill-posed linear inverse problem, because the measurement matrix $\mathbf{A}$ is singular and/or very ill conditioned. Since there is no unique solution for the underdetermined system in (3.23), additional information about $\mathbf{x}$ is typically employed in the form of regularization to stabilize and guide the search towards relevant solutions. MRI images are naturally *compressible* in an appropriate transform domain (such as wavelet or related sparsyfing transform) [Lustig 07], meaning that their sorted transform coefficients exhibit a power-law decay

(3.2) [Cevher 09a]. In the following we denote transform coefficients as $\boldsymbol{\theta} = \mathbf{P}\mathbf{x} \in \mathbb{C}^D$ where $\mathbf{P} \in \mathbb{C}^{D \times N}$ is the sparsyfing transform.

A common approach to estimate $\mathbf{x}$ given the measurements $\mathbf{y}$ in (3.23), is by solving the following constrained optimization problem [Lustig 07, Afonso 11]:

$$\min_{\mathbf{x} \in \mathbb{C}^N} \phi(\mathbf{P}\mathbf{x}) \quad \text{subject to} \quad \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \leq \epsilon \tag{3.24}$$

where $\phi : \mathbb{C}^N \mapsto \mathbb{R} \cup \{-\infty, +\infty\}$ is a regularization function and $\epsilon \geq 0$ is a parameter related to the noise variance which controls the fidelity of the reconstruction to the measured data. Choosing $\phi$ as $\ell_1$ norm: $\phi(\mathbf{x}) = \|\mathbf{x}\|_1$ leads to the basis pursuit denoising problem [Chen 98]. The so-called basis pursuit (BP) problem is the particular case of (3.24) for $\epsilon = 0$. In CS-MRI, $\phi$ is typically the $\ell_1$ norm, total variation (TV) norm in the image domain ($\mathbf{P} = \mathbf{I}$) or a linear combination of these two [Lustig 07, Lustig 08, Huang 11b, Ma 08]. Various reported methods focus on different aspects of this problem, such as improved iterative solvers [Ma 08,Yang 10], the use of efficient sparsifying transforms such as shearlets and curvelets [Aelterman 11,Smith 13,Qu 10,Pejoski 15,Liu 16] or trained dictionaries [Ravishankar 11, Qu 14, Huang 14], and adaptive sampling schemes [Bastounis 15, Adcock 14, Adcock 13]. Recent CS-MRI methods typically employ iterative reconstruction algorithms, both greedy and optimization-based. Well-known greedy methods include compressive sampling matching pursuit (CoSaMP) and subspace pursuit (SP) [Needell 09b, Baraniuk 10], iterative hard thresholding (IHT) [Blumensath 09] and its extensions [Blumensath 10], [Blumensath 12], [Rajani 12]. Among the methods that employ extensions of wavelets as sparsifying transforms are [Qu 10, Aelterman 11, Pejoski 15].

Methods employing convex non-smooth regularizers (TV and $\ell_1$) typically consider, instead of the original problem in (3.23), its equivalent unconstrained formulation [Afonso 11]:

$$\min_{\mathbf{x} \in \mathbb{C}^N} \phi(\mathbf{P}\mathbf{x}) + \frac{\mu}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \tag{3.25}$$

with $\mu > 0$. Many state-of-the-art methods for solving this problem belong to the iterative soft-thresholding (IST) [Daubechies 04] algorithms and their variants TwIST [Bioucas-Dias 07], FISTA [Beck 09], and SpaRSA [Wright 09]. The solution of (3.25) is usually defined in terms of the Moreau proximal mapping of $\phi$ [Combettes 05]

$$\boldsymbol{\Psi}_\phi(\mathbf{u}; \mu) = \underset{\mathbf{x} \in \mathbb{C}^N}{\operatorname{argmin}}\, \phi(\mathbf{x}) + \frac{\mu}{2}\|\mathbf{x} - \mathbf{u}\|_2^2 \tag{3.26}$$

For $\phi(\mathbf{x}) = \|\mathbf{x}\|_1$, this operator is component-wise soft-thresholding $\mathbf{\Psi}_{\ell_1}(\mathbf{u};\mu) = \text{soft}(\mathbf{u}, 1/\mu)$, which replaces each component of $\mathbf{u}$ by $sign(u)\max\{|u| - 1/\mu, 0\}$. For the TV norm: $\|\mathbf{x}\|_{\text{TV}}$, defined in (2.33), $\mathbf{\Psi}_{\text{TV}}(\mathbf{u};\mu)$ is computed using Chambolle's algorithm [Chambolle 04]. A recent method pFISTA [Liu 16] approximately solves the problem in (3.25) with a usage of tight frames such as contourlets, shift-invariant discrete wavelet (SIDWT) and patch based directional wavelet (PBDW) and $\ell_1$ norm regularization. Other recent approaches employ dictionary learning [Ravishankar 11, Huang 14] or patch-based nonlocal operators (PANO) [Qu 14]. Utilization of the AMP framework in MRI reconstruction is considered in [Eksioglu 18] where authors proposed extension of denoising-AMP (D-AMP) from [Metzler 16] with block matching 3D (BM3D) approach as a denoiser [Danielyan 11]. Although the behavior of AMP under the general random measurement matrices is still an open problem [Rangan 16], recent works suggest that subsampled unitary matrices are rather well suited for use in AMP [Ma 17].This was one of authors justification for developing BM3D-AMP-MRI approach [Eksioglu 18].

In the following, we review some of the representatives of state-of-the-art methods for MR reconstruction from undersampled data.

### 3.4.1 Nonlinear conjugate gradient iterative algorithm for $\ell_1$ penalized reconstruction

In a seminal paper on sparse MRI [Lustig 07], the unconstrained optimization problem in (3.25) was considered with the regularization function $\phi(\mathbf{Px}) = \|\mathbf{Px}\|_1$:

$$\min_{\mathbf{x}\in\mathbb{C}^N} \|\mathbf{Ax} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{Px}\|_1 \tag{3.27}$$

where $\lambda$ denotes the regularization parameter which determines the trade-off between the data consistency and the signal transform sparsity enforced by the regularization term. An iterative algorithm for solving (3.27) from [Lustig 07] performs in each iteration thresholding and cancellation of incoherent interference caused by measurements undersampling. This approach is closely related to classical solvers from [Donoho 06c, Rudin 92, Elad 07] with the difference that it employs nonlinear conjugate-gradient descent step in optimization procedure. It can be shown that a solution of (3.27) for which $\|\mathbf{Ax} - \mathbf{y}\|_2 \leq \epsilon$ is exactly the solution of (3.24) regardless of the chosen $\lambda$ (the value of $\lambda$

will influence the number of iterations needed to reach it). The authors in [Lustig 07] proposed solving (3.27) using a nonlinear conjugate gradient (CG) descent algorithm with backtracking line search. They approximate a gradient of the objective function in (3.27). Since the $\ell_1$ norm is not differentiable they involve the following approximation for the absolute function $|x| \approx \sqrt{x^2 + \mu}$, where $\mu$ is a positive smoothing parameter. With this approximation, a gradient of the objective function can be calculated. Then CG iterations are performed with a smoothing factor $\mu \in [10^{15}, 10^{-6}]$ until some stopping criteria is satisfied.

### 3.4.2 Augmented Lagrangian (AL) and alternating directional method of multipliers (ADMM)

In this subsection, we review some basic principles from optimization theory, which also form the basis for the CSALSA method that will be introduced in the following, and for the methods developed in this thesis.

Consider an unconstrained optimization problem of the form:

$$\min_{\mathbf{u}\in\mathbb{R}^N} f_1(\mathbf{u}) + f_2(g(\mathbf{u})) \tag{3.28}$$

where one of the constituent functions is a composition of two functions $f_2(g(\mathbf{u}))$ with $g : \mathbb{R}^N \mapsto \mathbb{R}^D$. A variable splitting (VS) technique creates a new variable, say $\mathbf{v}$, to serve as the argument of $f_2$, under the constraint that $g(\mathbf{u}) = \mathbf{v}$. With this, the unconstrained problem in (3.28) is transformed into a constrained problem:

$$\min_{\mathbf{u}\in\mathbb{R}^N, \mathbf{v}\in\mathbb{R}^D} f_1(\mathbf{u}) + f_2(\mathbf{v}) \quad \text{subject to} \quad g(\mathbf{u}) = \mathbf{v} \tag{3.29}$$

which is equivalent to the problem in (3.28) in the feasible set $\{(\mathbf{u}, \mathbf{v}) : g(\mathbf{u}) = \mathbf{v}\}$. The rationale behind variable splitting is that it may be easier to solve the constrained problem (3.29) than it is to solve its unconstrained counterpart in (3.28). VS is used in optimization which considers sum of two nonlinear monotone operators as objective function [Douglas 56, Lions 79, Eckstein 92]. Application of VS in linear inverse problems can be found in [Wang 08] where it is used to obtain a fast algorithm for TV-based restoration and in [Bioucas-Dias 08] to handle problems where instead of a single regularizer $\lambda\phi(\mathbf{x})$, we have a linear combination of two (or more) regularizers $\lambda_1\phi_1(\mathbf{x}) + \lambda_2\phi_2(\mathbf{x})$. The approach in [Wang 08, Bioucas-Dias 08] solves (3.29) by transforming it to an unconstrained version, with a quadratic penalty:

$$\min_{\mathbf{u}\in\mathbb{R}^N, \mathbf{v}\in\mathbb{R}^D} f_1(\mathbf{u}) + f_2(\mathbf{v}) + \frac{\alpha}{2}\|g(\mathbf{u}) - \mathbf{v}\|_2^2. \tag{3.30}$$

The problem in (3.30) is solved by alternating minimization with respect to $\mathbf{u}$ and $\mathbf{v}$, while $\alpha$ is slowly increasing to large values (a continuation process) in order to force the solution of (3.30) to approach that of (3.29) which in turn is equivalent to (3.28). The motivation behind the procedure described above i.e. transforming an unconstrained problem to a constrained one using variable splitting and then transforming again to another unconstrained formulation is that each step of the alternating minimization of (3.30) may be much easier than the original unconstrained problem in (3.28). Split-Bregman methods from [Bregman 67, Osher 05, Goldstein 09] utilize a similar variable splitting approach. Instead of using a quadratic penalty technique, those methods solve the constrained problem directly using a Bregman iterative algorithm [Osher 05,Xu 07,Yin 08]. Moreover, when a $g$ is a linear function i.e. $g(\mathbf{u}) = \mathbf{G}\mathbf{u}$ the Bregman iterative algorithm is equivalent to the augmented Lagrangian method [Esser 09,Setzer 09,Yin 08] which we review in the following.

Consider the constrained optimization problem

$$\min_{\mathbf{z}\in\mathbb{R}^N} E(\mathbf{z}) \quad \text{subject to} \quad \mathbf{H}\mathbf{z} - \mathbf{b} = 0 \tag{3.31}$$

where $\mathbf{b} \in \mathbb{R}^L$ and $\mathbf{H} \in \mathbb{R}^{L\times N}$, i.e. there are $L$ linear equality constraints. The augmented Lagrangian function for the problem in (3.31) is defined as:

$$\mathcal{L}_{\mathbf{H}}(\mathbf{z}, \boldsymbol{\lambda}, \mu) = E(\mathbf{z}) + \boldsymbol{\lambda}^T(\mathbf{b} - \mathbf{H}\mathbf{z}) + \frac{\mu}{2}\|\mathbf{H}\mathbf{z} - \mathbf{b}\|_2^2 \tag{3.32}$$

where $\boldsymbol{\lambda} \in \mathbb{R}^L$ is a vector of Lagrange multipliers and $\mu \geq 0$ is called the penalty parameter [Nocedal 06]. The so-called augmented Lagrangian method (ALM) [Nocedal 06], also known as the method of multipliers (MM) [Hestenes 69,Nocedal 06], iterates between minimizing $\mathcal{L}_{\mathbf{H}}(\mathbf{z}, \boldsymbol{\lambda}, \mu)$ with respect to $\mathbf{z}$, keeping $\boldsymbol{\lambda}$ fixed, and updating $\boldsymbol{\lambda}$, until some convergence criterion is satisfied. Using a complete-the-squares procedure, the terms added to $E(\mathbf{z})$ in (3.32) can be written as a single quadratic term (plus a constant independent of $\mathbf{z}$, thus irrelevant for minimization). This results in the form of the ALM algorithm given in Algorithm 5, which makes clear its equivalence with the Bregman iterative algorithm [Yin 08] where the variable $\mathbf{d}$ takes the role of Lagrange multipliers.

The introduced ALM method can be used to address the problem in (3.29) for the particular case when $g(\mathbf{u}) = \mathbf{G}\mathbf{u}$. Then the problem in (3.29) can be written in the form of (3.31) using the following definitions:

$$\mathbf{z} = \begin{bmatrix}\mathbf{u}\\ \mathbf{v}\end{bmatrix}, \quad \mathbf{b} = \mathbf{0}, \quad \mathbf{H} = \begin{bmatrix}\mathbf{G} - \mathbf{I}\end{bmatrix} \tag{3.33}$$

**Algorithm 5** ALM

**Input:** $k = 0, \mu > 0, \mathbf{d}^{\{0\}}$

1: **repeat**
2: $\quad \mathbf{z}^{\{k+1\}} = \underset{\mathbf{z}}{\operatorname{argmin}}\, E(\mathbf{z}) + \frac{\mu}{2}\|\mathbf{Hz} - \mathbf{d}^{\{k\}}\|_2^2$
3: $\quad \mathbf{d}^{\{k+1\}} = \mathbf{d}^{\{k\}} - (\mathbf{Hz}^{\{k+1\}} - \mathbf{b})$
4: $\quad k = k + 1$
5: **until** stopping criterion is satisfied

and

$$E(\mathbf{z}) = f_1(\mathbf{u}) + f_2(\mathbf{v}). \tag{3.34}$$

With these definitions, steps 2 and 3 of Algorithm 5 can be written as follows:

$$\begin{aligned}(\mathbf{u}^{\{k+1\}}, \mathbf{v}^{\{k+1\}}) &= \underset{\mathbf{u},\mathbf{v}}{\operatorname{argmin}}\, f_1(\mathbf{u}) + f_2(\mathbf{v}) + \frac{\mu}{2}\|\mathbf{Gu} - \mathbf{v} - \mathbf{d}^{\{k\}}\|_2^2 \\ \mathbf{d}^{\{k+1\}} &= \mathbf{d}^{\{k\}} - (\mathbf{Gu}^{\{k+1\}} - \mathbf{v}^{\{k+1\}}).\end{aligned} \tag{3.35}$$

The minimization problem given in (3.35) is non-trivial and in general can involve nonseparable quadratic, and possibly nonsmooth terms. A natural way to address the minimization in (3.35) is to use a nonlinear block-Gauss-Seidel (NLBGS) technique, which solves the problem in (3.35) by alternatingly minimizing it with respect to $\mathbf{u}$ and $\mathbf{v}$, while keeping the other variable fixed. Authors in [Goldstein 09] provided an experimental evidence that, before updating the variable $\mathbf{d}^{\{k\}}$, one iteration of NLBGS is enough to obtain efficient algorithm for solving the problem in (3.35). It turns out that the resulting algorithm is the so-called alternating direction method of multipliers (ADMM) [Eckstein 92] which we outlined in Algorithm 6. For comprehensive reviews of ALM, ADMM, and their relationship with Bregman and split-Bregman methods, we refer the reader to the following references [Setzer 09, Esser 09, Afonso 10, Afonso 11].

In the following subsection, we involve previously defined techniques (variable splitting, AL and ADMM) for deriving a constrained split augmented Lagrangian method (CSALSA).

### 3.4.3 Constrained split augmented Lagrangian method (CSALSA)

Although the formulation in (3.25) is equivalent to (3.24) for the appropriate $\mu$, and usually easier to solve, the formulation (3.24) has an

**Algorithm 6** ADMM

**Input:** $k = 0, \mu > 0, \mathbf{v}^{\{0\}}, \mathbf{d}^{\{0\}}$

1: **repeat**
2: $\quad \mathbf{u}^{\{k+1\}} = \underset{\mathbf{u}}{\operatorname{argmin}}\, f_1(\mathbf{u}) + \frac{\mu}{2}\|\mathbf{G}\mathbf{u} - \mathbf{v}^{\{k\}} - \mathbf{d}^{\{k\}}\|_2^2$
3: $\quad \mathbf{v}^{\{k+1\}} = \underset{\mathbf{v}}{\operatorname{argmin}}\, f_2(\mathbf{v}) + \frac{\mu}{2}\|\mathbf{G}\mathbf{u}^{\{k+1\}} - \mathbf{v} - \mathbf{d}^{\{k\}}\|_2^2$
4: $\quad \mathbf{d}^{\{k+1\}} = \mathbf{d}^{\{k\}} - (\mathbf{G}\mathbf{u}^{\{k+1\}} - \mathbf{v}^{\{k+1\}})$
5: $\quad k = k + 1$
6: **until** stopping criterion is satisfied

important advantage: the parameter $\epsilon$ in (3.24) has a clear interpretation in terms of the noise level, while setting the correct $\mu$ in (3.25) is not evident in practice and requires a clever algorithm to adjust it properly. Motivated by this, the authors in [Afonso 11] proposed an efficient algorithm for solving the constrained problem (3.24) directly. They incorporate the constraint from (3.24) in the objective function through the indicator function of the feasible set: the ellipsoid $E(\epsilon, \mathbf{A}, \mathbf{y}) = \{\mathbf{x} \in \mathbb{C}^N : \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon\}$. When using $\ell_1$ norm as the regularization function $\phi$, (3.24) gets transformed to the following problem:

$$\min_{\mathbf{x}\in\mathbb{C}^N} \|\mathbf{P}\mathbf{x}\|_1 + \iota_{E(\epsilon,\mathbf{I},\mathbf{y})}(\mathbf{A}\mathbf{x}) \tag{3.36}$$

where indicator function $\iota_{E(\epsilon,\mathbf{I},\mathbf{y})}$ returns 0 if its argument belongs to the set defined by the ellipsoid $E(\epsilon, \mathbf{A}, \mathbf{y})$ or $+\infty$ otherwise. The resulting unconstrained problem is further transformed into a different constrained problem by variable splitting [Courant 94] as was explained in the previous subsection. Then the obtained constrained problem is dealt with the alternating directional method of multipliers (ADMM) [Eckstein 92, Gabay 75, Glowinski 75] which belongs to the family of augmented Lagrangian (AL) techniques [Nocedal 06]. In this way the objective function is decoupled into a set of independent optimization subproblems which number depends on the number of terms in original objective function.

A variant of ADMM that minimizes a sum of more than two functions was derived in [Afonso 11]. This generalization of ADMM served as a basis for solving the problem in (3.36). Therefore, consider the more general case where there are $J$ functions involved in the mini-

mization, that is:

$$\min_{\mathbf{u}\in\mathbb{R}^D} \sum_{j=1}^{J} g_j(\mathbf{H}^{(j)}\mathbf{u}) \tag{3.37}$$

where $g_j : \mathbb{R}^{L_j} \mapsto \mathbb{R}$ are closed, proper convex functions, and $\mathbf{H}^{(j)} \in \mathbb{R}^{L_j \times D}$ are arbitrary matrices. If we make analogy with the derivation of ADMM explained in the previous subsection by connecting the terms in (3.35) and (3.37) in the following way:

$$\begin{aligned} f_1(\mathbf{u}) &= 0 \\ f_2(\mathbf{v}) &= \sum_{j=1}^{J} g_j(\mathbf{v}^j) \end{aligned} \tag{3.38}$$

where $\mathbf{v}^j \in \mathbb{R}^{L_j}$, $L = L_1 + \cdots + L_J$ and

$$\mathbf{v}_a^{\{k\}} = \begin{bmatrix} \mathbf{v}^{(1)} \\ \vdots \\ \mathbf{v}^{(J)} \end{bmatrix}^{\{k\}}, \quad \mathbf{d}_a^{\{k\}} = \begin{bmatrix} \mathbf{d}^{(1)} \\ \vdots \\ \mathbf{d}^{(J)} \end{bmatrix}^{\{k\}}, \quad \mathbf{G}_a = \begin{bmatrix} \mathbf{H}^{(1)} \\ \vdots \\ \mathbf{H}^{(J)} \end{bmatrix} \in \mathbb{R}^{L\times D} \tag{3.39}$$

then the Algorithm 6 can be applied for solving the problem in (3.37). Considering $f_1(\mathbf{u}) = 0$ turns step 3 of Algorithm 6 in a simple quadratic minimization problem which has a unique solution if $\mathbf{G}$ has a full column rank:

$$\begin{aligned} \mathbf{u}^{\{k+1\}} &= \underset{\mathbf{u}\in\mathbb{R}^D}{\operatorname{argmin}} \|\mathbf{G}_a\mathbf{u} - \mathbf{v}_a^{\{k\}} - \mathbf{d}_a^{\{k\}}\|_2^2 \\ &= (\mathbf{G}_a^H\mathbf{G}_a)^{-1}\mathbf{G}_a^H(\mathbf{v}_a^{\{k\}} + \mathbf{d}_a^{\{k\}}) \\ &= \Big[\sum_{j=1}^{J}(\mathbf{H}^{(j)})^H\mathbf{H}^{(j)}\Big]^{-1}\sum_{j=1}^{J}(\mathbf{H}^{(j)})^H((\mathbf{v}^{(j)})^{\{k\}} + (\mathbf{d}^{(j)})^{\{k\}}) \end{aligned} \tag{3.40}$$

Furthermore, step 4 of Algorithm 6 can be written in the following way:

$$\mathbf{v}_a^{\{k+1\}} = \underset{\mathbf{v}^{(1)},\ldots,\mathbf{v}^{(J)}}{\operatorname{argmin}} \; g_1(\mathbf{v}^{(1)}) + \cdots + g_J(\mathbf{v}^{(J)}) + \frac{\mu}{2}\|\mathbf{G}_a\mathbf{u}^{\{k+1\}} - \mathbf{v}_a - \mathbf{d}_a^{\{k\}}\|_2^2 \tag{3.41}$$

which allows decoupling the minimization in (3.41) into a set of $J$ independent ones:

$$(\mathbf{v}^{(j)})^{\{k+1\}} = \underset{\mathbf{v}\in\mathbb{R}^{L_j}}{\operatorname{argmin}}\, g_j(\mathbf{v}) + \frac{\mu}{2}\|\mathbf{v} - (\mathbf{s}^{(j)})^{\{k\}}\|_2^2 \tag{3.42}$$

**Algorithm 7** ADMM-2

**Input:** $k = 0, \mu > 0, (\mathbf{v}^{(1)})^{\{0\}}, \cdots, (\mathbf{v}^{(J)})^{\{0\}}, (\mathbf{d}^{(1)})^{\{0\}}, \cdots, (\mathbf{d}^{(J)})^{\{0\}}$

1: **repeat**
2: **for** $j = 1, \cdots, J$ **do**
3: $(\mathbf{r}^{(j)})^{\{k\}} = \sum_{j=1}^{J} (\mathbf{H}^{(j)})^H (\mathbf{v}^{(j)})^{\{k\}} + (\mathbf{d}^{(j)})^{\{k\}}$
4: **end for**
5: $\mathbf{u}^{\{k+1\}} = \left[\sum_{j=1}^{J} (\mathbf{H}^{(j)})^H \mathbf{H}^{(j)}\right]^{-1} (\mathbf{r}^{(j)})^{\{k\}}$
6: **for** $j = 1, \cdots, J$ **do**
7: $(\mathbf{v}^{(j)})^{\{k+1\}} = \mathbf{\Psi}_{g_j}(\mathbf{H}^{(j)}\mathbf{u}^{\{k+1\}} - (\mathbf{d}^{(j)})^{\{k\}}; \mu)$
8: $(\mathbf{d}^{(j)})^{\{k+1\}} = (\mathbf{d}^{(j)})^{\{k\}} - \mathbf{H}^{(j)}\mathbf{u}^{\{k+1\}} + (\mathbf{v}^{(j)})^{\{k+1\}}$
9: **end for**
10: $k = k + 1$
11: **until** stopping criterion is satisfied

where $j = 1, 2, ..., J$ and $(\mathbf{s}^{(j)})^{\{k\}} = \mathbf{H}^{(j)}\mathbf{u}^{\{k+1\}} - (\mathbf{d}^{(j)})^{\{k\}}$. Notice that the update equations in (3.42) has the the form of the Moreau proximal map from (3.26) and can be written as:

$$(\mathbf{v}^{(j)})^{\{k+1\}} = \mathbf{\Psi}_{g_j}((\mathbf{s}^{(j)})^{\{k\}}; \mu) \tag{3.43}$$

The obtained generalized version of ADMM algorithm (herein referred to as ADMM-2) is outlined in Algorithm 7. Then, the problem in (3.36) can be mapped to the form in (3.37) with $J = 2$ and using the following correspondences:

$$\begin{aligned} g_1 &= \|\mathbf{P}\mathbf{x}\|_1 \\ g_2 &= \iota_{E(\epsilon, \mathbf{I}, \mathbf{y})} \\ \mathbf{H}^{(1)} &= \mathbf{P} \\ \mathbf{H}^{(2)} &= \mathbf{A} \end{aligned} \tag{3.44}$$

In order to simplify the notation we use the following relations $\mathbf{w}^{\{k\}} = (\mathbf{v}^{(1)})^{\{k\}}, \mathbf{v}^{\{k\}} = (\mathbf{v}^{(2)})^{\{k\}}, \mathbf{c}^{\{k\}} = (\mathbf{d}^{(1)})^{\{k\}}, \mathbf{b}^{\{k\}} = (\mathbf{d}^{(2)})^{\{k\}}$ and define regularization term as $\phi(\mathbf{x}) = \|\mathbf{P}\mathbf{x}\|_1$. Introduced relations together with correspondences in (3.44), result in a constrained split augmented Lagrangian (CSALSA) algorithm which is a special case of Algorithm 7. The CSALSA method is outlined Algorithm 8 with the simplified notation.

In [Afonso 11], the authors assume that the sparsifying transform $\mathbf{P}$ is the analysis operator of a 1-tight (Parseval) frame, thus

**Algorithm 8** CSALSA

**Input:** $k = 0, \mu > 0, \mathbf{w}^{\{0\}}, \mathbf{v}^{\{0\}}, \mathbf{c}^{\{0\}}, \mathbf{b}^{\{0\}}$

1: **repeat**
2: $\quad \mathbf{r}^{\{k\}} = \mathbf{P}^H(\mathbf{w}^{\{k\}} + \mathbf{c}^{\{k\}}) + \mathbf{A}^H(\mathbf{v}^{\{k\}} + \mathbf{b}^{\{k\}})$
3: $\quad \mathbf{x}^{\{k+1\}} = (\mathbf{P}^H\mathbf{P} + \mathbf{A}^H\mathbf{A})^{-1}\mathbf{r}^{\{k\}}$
4: $\quad \mathbf{w}^{\{k+1\}} = \mathbf{\Psi}_{\phi(\mathbf{x})}(\mathbf{P}\mathbf{x}^{\{k+1\}} - \mathbf{c}^{\{k\}}; \mu)$
5: $\quad \mathbf{v}^{\{k+1\}} = \mathbf{\Psi}_{\iota_{E(\epsilon,\mathbf{I},\mathbf{y})}}(\mathbf{A}\mathbf{x}^{\{k+1\}} - \mathbf{b}^{\{k\}}; \mu)$
6: $\quad \mathbf{c}^{\{k+1\}} = \mathbf{c}^{\{k\}} - \mathbf{P}\mathbf{u}^{\{k+1\}} + \mathbf{w}^{\{k+1\}}$
7: $\quad \mathbf{b}^{\{k+1\}} = \mathbf{b}^{\{k\}} - \mathbf{A}\mathbf{u}^{\{k+1\}} + \mathbf{v}^{\{k+1\}}$
8: $\quad k = k + 1$
9: **until** stopping criterion is satisfied

$\mathbf{P}^H\mathbf{P} = \mathbf{I}$ which simplifies the step 3 of CSALSA to:

$$\mathbf{x}^{\{k+1\}} = (\mathbf{I} + \mathbf{A}^H\mathbf{A})^{-1}\mathbf{r}^{\{k\}}. \tag{3.45}$$

The measurement matrix $\mathbf{A}$ can be written as $\mathbf{A} = \mathbf{SF}$, where $\mathbf{S}$ is a $M \times N$ binary sampling matrix with $M < N$ formed by a subset of rows of the identity matrix chosen according to the sampling trajectory, and $\mathbf{F}$ is a DFT matrix. With the fact that matrix $\mathbf{S}$ satisfies $\mathbf{S}\mathbf{S}^H = \mathbf{I}$ and using the Sherman-Morrison-Woodbury (SMW) matrix inversion lemma [Deng 11], the matrix inversion in (3.45) can be computed as:

$$(\mathbf{I} + \mathbf{A}^H\mathbf{A})^{-1} = \mathbf{I} - \frac{1}{2}\mathbf{F}^H\mathbf{S}^H\mathbf{S}\mathbf{F} \tag{3.46}$$

where $\mathbf{S}^H\mathbf{S}$ has the structure of identity matrix with some zeros on the diagonal determined by the undersampling pattern. Then the computational cost of CSALSA step 4 is defined by the products of $\mathbf{F}$ and $\mathbf{F}^H$, which are implemented by fast Fourier transform (FFT), and hence is equal to $\mathcal{O}(N \log N)$. The CSALSA method has proved excellent performance in MRI-reconstruction and besides that in other linear inverse imaging problems such as image inpainting, image debluring and denoising. The CSALSA framework would be one of the important parts in building reconstruction methods in this thesis. A detail derivation of our methods based on CSALSA framework will be given in Chapter 4.

### 3.4.4 Fast composite splitting algorithm (FCSA)

When the regularization function $\phi$ (3.25) involves more than one term, a suitable solver is ADMM algorithm or its particular version CSALSA that was described in the previous subsection. In

**Algorithm 9** FISTA

**Input:** $k = 1, \rho = 1/L_f, \mathbf{r}^{\{1\}} = \mathbf{x}^{\{0\}}, t^{\{1\}} = 1$

1: **repeat**
2: $\quad \mathbf{x}_g = \mathbf{r}^{\{k\}} - \rho \nabla f(\mathbf{r}^{\{k\}})$
3: $\quad \mathbf{x}^{\{k\}} = \mathbf{\Psi}_{g(\mathbf{x})}(\mathbf{x}_g; \rho)$
4: $\quad t^{\{k+1\}} = \frac{1+\sqrt{1+4(t^{\{k\}})^2}}{2}$
5: $\quad \mathbf{r}^{\{k+1\}} = \mathbf{x}^{\{k\}} + \frac{t^{\{k\}}-1}{t^{\{k+1\}}}(\mathbf{x}^{\{k\}} - \mathbf{x}^{\{k-1\}})$
6: $\quad k = k + 1$
7: **until** stopping criterion is satisfied

[Huang 11a, Huang 11b], the authors consider an approach which results in a different iterative framework named fast composite splitting algorithm (FCSA). The FCSA method minimizes linear combination of three terms corresponding to a least square data fitting, total variation (TV) and $\ell_1$ norm regularization:

$$\min_{\mathbf{x}\in\mathbb{C}^N} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \alpha\|\mathbf{x}\|_{\mathrm{TV}} + \beta\|\mathbf{P}\mathbf{x}\|_1 \tag{3.47}$$

in order to recover the MR image $\mathbf{x}$. Based on the fast iterative shrinkage-thresholding algorithm (FISTA) from [Beck 09] the FCSA method proposed efficient solution for the composite $\ell_1$ and TV regularization in (3.47).

The FISTA considers to minimize the following optimization problems

$$\min_{\mathbf{x}} f(\mathbf{x}) + g(\mathbf{x}) \tag{3.48}$$

where objective function contains one smooth convex function $f(\mathbf{x})$ with Lipschitz constant $L_f$ and convex function $g(\mathbf{x})$ which may be nonsmooth. Two main steps in the FISTA algorithm are: i) gradient decent step, which involves calculation of gradient of the function $f(\mathbf{x})$, and ii) proximal map step $\mathbf{\Psi}_{g(\mathbf{x})}(\mathbf{x}; \rho)$, given a continuous convex function $g(\mathbf{x})$ and arbitrary scalar $\rho > 0$. FISTA steps are given in Algorithm 9.

However, no efficient algorithm exists to solve the proximal map which involves composite regularization function $g(\mathbf{x}) = \alpha\|\mathbf{x}\|_{\mathrm{TV}} + \beta\|\mathbf{P}\mathbf{x}\|_1$ given in (3.47) i.e.

$$\mathbf{x}^{\{k\}} = \mathbf{\Psi}_{\alpha\|\mathbf{x}\|_{\mathrm{TV}}+\beta\|\mathbf{P}\mathbf{x}\|_1}(\mathbf{x}_g; \rho). \tag{3.49}$$

In order to solve problem in (3.49), authors in [Huang 11b] proposed involving composite splitting technique used in composite splitting denoising (CSD) method from [Huang 11a]. With this technique a solution for

**Algorithm 10** FCSA

**Input:** $k = 1, \rho = 1/L_f, \alpha, \beta, \mathbf{r}^{\{1\}} = \mathbf{x}^{\{0\}}, t^{\{1\}} = 1$

1: **repeat**
2: $\quad \mathbf{x}_g = \mathbf{r}^{\{k\}} - \rho \nabla f(\mathbf{r}^{\{k\}})$
3: $\quad \mathbf{x}_1 = \mathbf{\Psi}_{2\alpha\|\mathbf{x}\|_{\mathrm{TV}}}(\mathbf{x}_g; \rho)$
4: $\quad \mathbf{x}_2 = \mathbf{\Psi}_{2\beta\|\mathbf{P}\mathbf{x}\|_1}(\mathbf{x}_g; \rho)$
5: $\quad \mathbf{x}^{\{k\}} = (\mathbf{x}_1 + \mathbf{x}_2)/2$
6: $\quad \mathbf{x}^{\{k\}} = \mathrm{project}(\mathbf{x}^{\{k\}}, [l, u])$
7: $\quad t^{\{k+1\}} = \frac{1+\sqrt{1+4(t^{\{k\}})^2}}{2}$
8: $\quad \mathbf{r}^{\{k+1\}} = \mathbf{x}^{\{k\}} + \frac{t^{\{k\}}-1}{t^{\{k+1\}}}(\mathbf{x}^{\{k\}} - \mathbf{x}^{\{k-1\}})$
9: $\quad k = k + 1$
10: **until** stopping criterion is satisfied

the proximal map problem in (3.49) is approximated as a linear combination of solutions of proximal maps related to each involved regularization function separately. Therefore, combining the CSD with FISTA a new algorithm named FCSA outlined in Algorithm 10 is proposed. The steps 3 and 4 of the algorithm correspond to independently performed proximal maps for $\ell_1$ and TV norm respectively. The average of the obtained solutions $\mathbf{x}_1$ and $\mathbf{x}_2$ from step 5 is further projected onto the range of $\mathbf{x}$ defined by $[l, u]$ in step 6. The project function maps all components of vector $\mathbf{x}$ to $l$ if they are less than $l$, to $u$ if they are greater than $u$ and keep them unchanged if they are in the range $[l, u]$. The proposed FCSA method has fast convergence properties and has been shown to significantly outperform the NLCG method introduced in subsection 3.4.1 and two state-of-the-art methods TVCMRI [Ma 08] and RecPF [Yang 10] in terms of both accuracy and complexity. In Chapter 3 we will use the FCSA framework without projection function in step 6, as a basis for the development of our methods.

### 3.4.5 Wavelet tree sparsity MRI (WaTMRI)

Recent work demonstrates benefits of encoding *structure* of the sparse, information-bearing coefficients, either in the acquisition [Adcock 14] or in the reconstruction [Deka 18, Panić 17a, Lai 16, Chen 14, Chen 12, Pižurica 11, He 10] stages. Subband coefficients of natural images, including MRI, obey certain structures, which can be viewed in terms of fixed groups (like wavelet trees) or statistically (certain clustering configuration are more likely than others).

One way to improve sparsity constraint is to encode structure among coefficients and replace the standard sparsity with tree sparsity which the most existing tree-based compressed sensing algorithms actually do, e.g. [He 10, He 09, Som 12, He 10]. In practice, however the wavelet coefficients cannot perfectly match the theoretical assumption of tree sparsity. In order to tackle this practical issue for MR images, the approach of [Chen 14] named Wavelet Tree Sparsity MRI (WaTMRI), approximates the tree structure with overlapping groups. They introduce a new regularization term in the objective function besides TV and $\ell_1$ regularization:

$$\min_{\mathbf{x}\in\mathbb{C}^N} \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2^2 + \alpha\|\mathbf{x}\|_{\mathrm{TV}} + \beta\left(\|\mathbf{P}\mathbf{x}\|_1 + \sum_{g\in\mathcal{G}}\|(\mathbf{P}\mathbf{x})_g\|_2\right) \tag{3.50}$$

which forces sparsity of the parent-child groups $g$ in the wavelet quadtree where $\mathcal{G}$ denotes all such groups in a wavelet tree. The objective function in (3.50) is decomposed into three sub-problems with fast composite splitting algorithm (FCSA) [Huang 11a]. Those sub-problems are further solved using proximal map operators given in (3.26). Incorporating TV and $\ell_1$ regularization next to the wavelet-tree sparsity term, make the method more robust on undersampled artifacts in MR images compared with previous tree structured models.

### 3.4.6 Lattice Split Bregman (LaSB)

In literature, much less attention has been devoted to modeling intra-scale coefficient dependencies, such as the spatial clustering of subband data in CS-MRI. The so-called Lattice Matching Pursuit (LaMP) algorithm of [Cevher 09b], utilized a Markov Random Field (MRF) model to encode structure in sparse images. The LaMP is a greedy algorithm, based on the orthogonal matching pursuit (OMP) approach. Developed for signals on 2D lattices (images) it includes the likelihood of the signal support which is iteratively evaluated and optimized under an Ising model as an example of MRF. With the involved Ising model, a partial knowledge of the sparse signal support greatly decreases the ambiguity and thus the size of the search space for the remaining unknown part, thus accelerating the algorithm. Although the LaMP has demonstrated great potentials in surveillance videos, a limited evaluation is shown for MRI reconstruction. In [Cevher 09b] only a synthetic Shepp-Logan phantom, which is sparse in the spatial domain $\mathbf{P} = \mathbf{I}$, is considered in evaluation thus without involving compressible MRI images.

A related structured-sparsity approach, but with important differences including a different likelihood model, the use of optimization framework instead of the greedy one and a different inference approach was proposed in [Pižurica 11]. To make a parallel with LaMP, the authors named this method LaSB as an acronym of 'Lattice Split-Bregman' and evaluated it thoroughly in MRI reconstruction.

The authors in [Pižurica 11] addressed the following fundamental problem. The existing optimization-based approaches (including Split-Bregman and related ADMM-based methods, like CSALSA, FISTA and FCSA described above) typically assumed that the image coefficients are statistically independent. Note that the $\ell_1$ norm $\phi(\mathbf{x}) = ||\mathbf{x}||_1$ assumes that the underlying coefficients are statistically independent and double exponential (Laplace) distributed. Under such statistical independence assumptions, Moreau proximal mapping of a general problem defined in (3.25) reduces to a soft-thresholding operator (or to hard-thresholding in the case of $||\mathbf{x}||_0$). The question is how this update rule should be changed to account for the statistical dependencies that exist among the image coefficients (e.g., in each subband of a wavelet or wavelet-like decomposition).

Without explicitly deriving such an update rule, the authors in [Pižurica 11] defined a lattice selector operation, which selects (in each iteration of the ADMM-based optimization algorithm) those coefficients that live on the estimated spatial support. The corresponding regularization function $\phi(\mathbf{x})$ accounts for the statistical dependencies among the transform coefficients that are exploited to estimate the likely spatial support, i.e., mask $\mathbf{l}$.

Formally, denoting the image coefficients by $\mathbf{d} = \mathbf{P}\mathbf{x}$, and the i-th subband coefficient by $\mathbf{d}_{[i]} = d_i$, the soft-thresholding update (that follows from the $\ell_1$ regularization can be written as

$$\mathbf{\Psi}_{\ell_1}(\mathbf{d}_{[i]}; \mu) = sign(\mathbf{d}_{[i]}) \max\{|\mathbf{d}_{[i]}| - 1/\mu, 0\} \tag{3.51}$$

and with the same notation, the lattice selector in LaSB is:

$$\mathbf{\Psi}_{\phi(\mathbf{d})}(\mathbf{d}_{[i]}; \hat{\mathbf{l}}_{[i]}) = \Lambda(\mathbf{d}_{[i]}, \hat{\mathbf{l}}_{[i]}), \tag{3.52}$$

with

$$\Lambda(d_i, \hat{l}_i) = \begin{cases} d_i, & \hat{l}_i = 1 \\ 0, & \text{otherwise.} \end{cases} \tag{3.53}$$

The mask $\mathbf{l}$ is calculated taking into account the statistical dependencies among the elements of $\mathbf{d}$. Observe that while soft-thresholding in (3.51)

operates on each element of $\mathbf{d}$ independently, the lattice selector in (3.52) takes into account the dependencies among the elements in $\mathbf{d}$ through the fact that $\mathbf{l}_{[i]} = l_i$ depends on all $d_j$ (and not only for $j = i$). Hence, the proximal map $\Lambda(d_i, \hat{l}_i)$ indirectly through the label mask $\hat{\mathbf{l}}$ propagates information about coefficient dependencies into a hard-threshold regularization on each image coefficients $d_i$ in (3.53). This way, LaSB incorporates into the reconstruction prior knowledge about the statistical dependencies of the sparse coefficients. Dependencies among image coefficients (e.g. shearlet coefficients) are modeled with an Ising model on 'hidden labels' $l_i \in \{-1, 1\}$ which are attached to the (shearlet) coefficients $d_i$. Estimation of the signal support represented with binary label mask $\mathbf{l}$ is done using a Metropolis-sampler inference algorithm. Notice that unlike CSALSA, LaSB used the $\ell_2$ norm (and thus not the indicator function defined in subsection 3.4.3) in the unconstrained formulation of the problem. This led to using few iterations of CG method for solving the minimization problem in step 2 of Algorithm 6. The initial results in [Pižurica 11] demonstrated great potential of MRF-based regularization for rapid MRI imaging, which deserves to be studied more deeply. Motivated by these encouraging results, we develop and evaluate thoroughly efficient MRF-based CS-MRI methods based on both optimization and greedy approaches.

### 3.4.7 Conclusion

In this chapter, we reviewed basic principles of the CS theory and its utilization in MR reconstruction from partial data i.e. CS-MRI. We explained the conditions under which CS can be successfully applied in the considered reconstruction problem. Among those we mostly concentrated on signal sparsity (i.e. sparse signal representation) and wavelet-like sparsifying transforms which will be employed in our algorithms in the following chapters. A detailed review of MR recovery methods was given with particular attention to those which will serve as a basis in developing our methods later on.

In the next chapter, we introduce a Markov Random Field (MRF) statistical framework which will be our important pillar in the construction of the statistical signal model. We will then incorporate the developed statistical signal model (MRF-based prior) into optimization solvers and greedy iterative framework in Chapters 4 and 5 and present improvements obtained in reconstruction.

# 4

# MRI reconstructions with MRF priors

*The mind uses its faculty for creativity only when experience forces it to do so.*
—Henri Poincaré

In this chapter, we derive our statistical signal model based on a Markov random field and propose its utilization in CS-MRI reconstruction. We start with a review of the theoretical background of Markov random field models and emphasize their expressiveness in spatial statistics modeling. Then we propose an MRF prior which serves as a signal model and encodes the spatial context of image coefficients. This model will be incorporated through regularization functions in the CS-MRI reconstruction. Under the optimization-based problem formulation, we derive four methods with the proposed regularization functions. They differ in view of the adopted iterative optimization frameworks, the way how the MRF-based regularization is effectively incorporated (alone or within a compound regularization) and in terms of the actual MRF model and its parameter specification. In particular, we derive two MRF-based reconstruction algorithms within the CSALSA framework: one with the MRF-based regularization alone and another in combination with TV regularization. Under an alternative composite splitting framework we derive another MRF-based approach, where we now also admit a more general (anisotropic) MRF prior and develop an automatic procedure for the parameter estimation. Two variants of this approach are analyzed (with and without acceleration).

## 4.1 Introduction to Markov Random Fields

Markov random fields (MRF) belong to a class of undirected probabilistic graphical models (PGMs). They are based on graph representation for encoding a distribution over a multi-dimensional space used for modeling context-dependent entities such as image pixels or correlated features extracted from image pixels. The joint distribution which is characterizing mutual influences among such entities is efficiently represented through MRF models using local interactions i.e. 'potential functions' or factors. The theorem which states the equivalence between MRF's and Gibbs distributions, established by Hammersley and Clifford in 1971 and further developed by Besag in [Besag 74], was of crucial importance for the usage of MRF models in practice [Li 09]. The MRF models also express the conditional independence structure between random variables. This property becomes very useful for analyzing the spatial or contextual dependencies of physical phenomena and will be utilized in this thesis to establish a probabilistic description of interactions among sparse image coefficients.

Before defining the MRF model, we need to introduce the notion of sites and the concept of neighborhood systems. Since MRF is used for modeling spatial data, the domain for MRF models is represented with the set of sites or locations which have some spatial connotation. The site denotes a point in the Euclidean space, such as the position of the image pixel in a finite two-dimensional rectangular grid (lattice) [Li 09, Winkler 12]. In this thesis we restrict our attention on images, as an instance of spatially generated data, and define the set of sites which denotes domain (rectangular lattice) for an image of size $K \times K$ as:

$$\mathcal{S} = \{(m, n) \mid 1 \leq m, n \leq K\}. \tag{4.1}$$

Therefore the elements of set $S$ correspond to the points at which image is sampled, i.e., to the location of image pixels, where $m$ and $n$ denote indices for a row and a column respectively. A single-number index for the site is also very common. For a $K \times K$ image, a pixel at position $(m, n)$ can be conveniently reindexed by a single number $i$, where $i$ takes on values in $1, 2, ..., d$ with $d = K \times K$.

A neighborhood system for sites in $\mathcal{S}$ represents interrelationship among them. Using single-number site index to simplify notation, neighborhood system for $\mathcal{S}$ is defined as:

$$\mathcal{N} = \{\mathcal{N}_i \mid \forall i \in \mathcal{S}\} \tag{4.2}$$

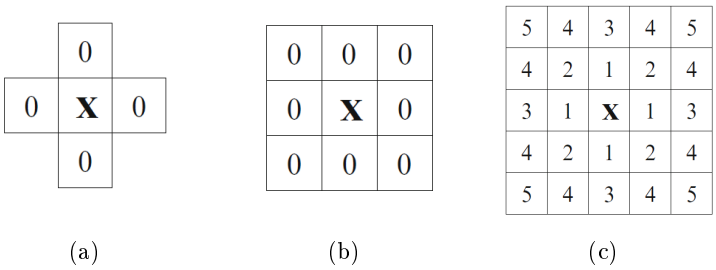



**Figure 4.1:** Neighborhood systems: $(a)$ $1^{st}$-order, $(b)$ $2^{nd}$-order, $(c)$ $5^{th}$-order

where $\mathcal{N}_i$ denotes the set of sites neighboring site $i$. The neighboring relationship has the following properties:

1. A site is not a neighbor of itself: $i \notin \mathcal{N}_i$.
2. The neighboring relationship is mutual if: $i \in \mathcal{N}_{i'} \iff i' \in \mathcal{N}_i$.

In the case of a regular lattice $\mathcal{S}$ the set of neighbors of $i$ is commonly defined as the set of sites within a radius $r$ from $i$:

$$\mathcal{N}_i = \{i' \in \mathcal{S} \mid \mathbf{dist}(p_i, p_{i'}) \leq r,\ i \neq i'\} \tag{4.3}$$

where $\mathbf{dist}(p_i, p_{i'})$ denotes the Euclidian distance between the pixel coordinates $p_i, p_{i'}$ and $r$ takes an integer value. It is worth mentioning that sites at or near the boundaries have fewer neighbors. In the first-order neighborhood system, also called the 4-neighborhood system, every (interior) site $(i,j)$ has four neighbors $\mathcal{N}_{i,j} = \{(i-1,j),(i+1,j),(i,j-1),(i,j+1)\}$, while in the second-order neighborhood system, also called the 8-neighborhood system, there are eight neighbors for every (interior) site $\mathcal{N}_{i,j} = \{(i-1,j),(i-1,j-1),(i-1,j+1),(i+1,j),(i+1,j-1),(i+1,j+1),(i,j-1),(i,j+1)\}$. Fig. 4.1 illustrates these neighborhood systems. The numbers $\{1,2,3,4,5\}$ shown in Fig. 4.1 (c) indicate the outermost neighboring sites in the $5^{\text{th}}$-order neighborhood system.

With definition of the notion of sites $\mathcal{S}$ and the neighborhood system $\mathcal{N}$ an undirected graph $\mathcal{G} = (\mathcal{S}, \mathcal{N})$ can be constructed, where $\mathcal{S}$ contains the nodes and $\mathcal{N}$ determines the connections between the graph nodes according to the neighboring relationship. Further a notion of *clique* $c$ is introduced for the graph $\mathcal{G}$ and defined as a subset of sites in $\mathcal{S}$ where all nodes (sites) are directly connected i.e. they are in mutual neighborhood relations. Clique consists of either a single-site $c = \{i\}$, a pair of neighboring sites $c = \{i, i'\}$, a triple of neighboring

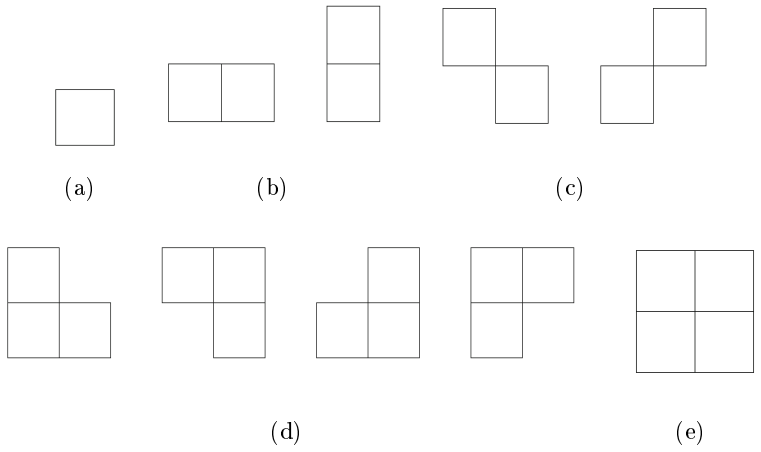



**Figure 4.2:** Cliques on a lattice of regular cites.

sites $c = \{i, i', i''\}$, and so on. If we denote the single-site, the pair-site, and the triple-site cliques with $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3$ respectively, then the collections of all cliques for graph $\mathcal{G}$ is

$$\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2 \cup \mathcal{C}_3 \cdots \tag{4.4}$$

where '$\cdots$' denotes possible sets of larger cliques. The clique types for $\mathcal{G}$ of a regular lattice are determined by its size, shape, and orientation. In Fig. 4.2(a)-(e) we show types for the neighborhood systems for a lattice. For the first-order neighborhood system only available cliques are single-site and horizontal and vertical pair-site cliques (a) and (b). The clique types for the second-order neighborhood systems besides those in (a) and (b) also include diagonal pair-site cliques (c) and triple-site (d) and quadruple-site (e) cliques. With the enlargment of the neighborhood system the number of cliques grows rapidly and so do the computational expenses involved.

With the previously introduced creation of graph with sites $\mathcal{S}$ and neighborhood system $\mathcal{N}$ we are ready to define a Markov Random Field. Let $\mathbf{F} = \{F_1, F_2, F_3, .., F_i, .., F_d\}$ be a family of random variables defined on the set $\mathcal{S}$ with the usage of single-number index notation. Each random variable $F_i$ takes a value $f_i$ from the label set $\mathcal{L}_i$ of possible realizations (hereafter 'labels') of random variable $F_i$ which may be continuous or discrete. The family $F$ is then called a random field. With the $F_i = f_i$ we denote the event that $F_i$ takes the value $f_i$ and the notation $(F_1 = f_1, ..F_i = f_i.., F_d = f_d)$ represents the joint event. We can use

abbreviation $F = f$ for the joint event where $f = \{f_1, f_2, .., f_i, .., f_d\}$ is a *configuration* of $F$ corresponding to the realization of the field. If a label set $\mathcal{L}_i$ only contains a discrete labels, the probability of the event $F_i = f_i$ is denoted as $P(F_i = f_i)$, abbreviated $P(f_i)$, and the joint probability is denoted $P(\mathbf{F} = \mathbf{f}) = P(F_1 = f_1, ..F_i = f_i.., F_d = f_d)$ and abbreviated $P(\mathbf{f})$. In the case of a continuous label set $\mathcal{L}_i$ we have a probability density functions (p.d.f.) $p(F_i = f_i)$ and $p(F = \mathbf{f})$. If all the sites have the same label set denoted as $\mathcal{L}$, which is usually the case, then the set of all possible labelings (that is, the configuration space) is the Cartesian product

$$\mathbb{F} = \underbrace{\mathcal{L} \times \mathcal{L} \cdots \times \mathcal{L}}_{\text{d times}} = \mathcal{L}^d \tag{4.5}$$

where $d$ is the total number of sites in $\mathcal{S}$. When the common label set for all sites is the real line namely $\mathcal{L} = \mathbb{R}$ then $\mathbb{F} = \mathbb{R}^d$ is the $d$-dimensional real space. If $\mathcal{L}$ is a discrete set then the size of $\mathbb{F}$ is combinatorial.

We say that $\mathbf{F}$ is a Markov random field on $\mathcal{S}$ with respect to (w.r.t.) a neighborhood system $\mathcal{N}$ if and only if the following two conditions are satisfied:

$$P(\mathbf{f}) > 0, \quad \forall \mathbf{f} \in \mathbb{F} \quad \text{(positivity)} \tag{4.6}$$

and

$$P(f_i \mid f_{\mathcal{S}-\{i\}}) = P(f_i \mid f_{\mathcal{N}_i}) \quad \text{(Markovianity)} \tag{4.7}$$

where $\mathcal{S} - \{i\}$ is the set difference containing all sites except site $i$, $f_{\mathcal{S}-\{i\}}$ denotes the set of labels at the sites in $\mathcal{S} - \{i\}$, and

$$f_{\mathcal{N}_i} = \{f_{i'} \mid i' \in \mathcal{N}_i\} \tag{4.8}$$

stands for the set of labels at the sites neighboring $i$. When the positivity condition is satisfied, the joint probability $P(\mathbf{f})$ of any random field is uniquely determined by its local conditional probabilities [Besag 74]. From the Markovianity property we see that the local characteristics of $\mathbf{F}$ are described through the neighborhood system $\mathcal{N}$ since only neighboring labels have direct interactions with each other. The MRF is said to be homogeneous if $P(f_i \mid f_{\mathcal{N}_i})$ is independent of the relative location of the site $i$ in $\mathcal{S}$.

Two approaches can be considered for specifying an MRF: one in terms of conditional probabilities $P(f_i \mid f_{\mathcal{N}_i})$ and the other in terms of the joint probabilities $P(\mathbf{f})$. Besag in his work [Besag 74] argued for the joint probability approach in view of the disadvantages of the conditional probability approach. Using the associated conditional probabilities there is no obvious method for deducing the joint probability.

Another thing is that conditional probabilities themselves are subject to some non-obvious and highly restrictive consistency conditions. The specification of the equilibrium of a statistical process is in terms of joint probability rather than the conditional distribution of the variables. With the theoretical results from Hammersley and Clifford in 1971 about the equivalence between MRF and Gibbs distribution, there is a mathematically tractable way for specifying the joint probability of the MRF. In the next subsection, we introduce the notion of Gibbs distribution and connect it to MRF.

### 4.1.1 Gibbs distribution and its equivalence with MRF

A Gibbs random field (GRF) is defined as a set of random variables $\mathbf{F}$ on a site set $\mathcal{S}$ w.r.t. $\mathcal{N}$ if and only if its configuration obey a Gibbs distribution which takes the form

$$P(\mathbf{f}) = Z^{-1} \times e^{-\frac{1}{T}H(\mathbf{f})} \tag{4.9}$$

where

$$Z = \sum_{\mathbf{f}\in\mathbb{F}} e^{-\frac{1}{T}H(\mathbf{f})} \tag{4.10}$$

is a normalizing constant called the partition function, $T$ is constant called the temperature and $H(\mathbf{f})$ is the energy function. The energy is represented as a sum of clique potentials (equation (4.11)), denoted as $V_c(\mathbf{f})$, over all possible cliques in a clique collection set $\mathcal{C}$.

$$H(\mathbf{f}) = \sum_{c\in\mathcal{C}} V_c(\mathbf{f}) \tag{4.11}$$

Local configuration of clique $c$ determines the value of clique potential $V_c(\mathbf{f})$. We say that GRF is homogeneous if $V_c(\mathbf{f})$ is independent of the relative position of the clique $c$ in $\mathcal{C}$ and that is isotropic if $V_c(\mathbf{f})$ is independent of the orientation of $c$. The calculation of Gibbs distribution depends on the evaluation of the partition function $Z$. Since $Z$ is the sum of all possible configuration in $\mathbb{F}$ and the case of a discrete $\mathcal{L}$ there is a combinatorial number of elements in $\mathbb{F}$, the evaluation of $Z$ is prohibitive even for problems of moderate size. The occurrence of a particular configuration or 'pattern' $\mathbf{f}$ is measured by the probability $P(\mathbf{f})$. Lower value of energy function $H(\mathbf{f})$ indicates that configuration $\mathbf{f}$ is more probable. The parameter $T$ controls the shape of the distribution in the following sense: when temperature $T$ is high all configurations

tend to be equally distributed while near the zero temperature the distribution concentrates around the global energy minimum. With this in mind, we can generate a class of 'patterns' by sampling the configuration space $\mathbb{F}$ according to the $P(\mathbf{f})$ which is completely determined by given temperature $T$ and energy function $H(\mathbf{f})$.

In case of a discrete labels a clique potential $V_c(\mathbf{f})$ can be specified by a number of parameters. For a triple clique $c = \{i, i', i''\}$ let $f_c = (f_i, f_{i'}, f_{i''})$ be its local configuration which has a finite number of states and therefore $V_c(\mathbf{f})$ takes a finite number o values. For a continuous labeling problem $V_c(\mathbf{f})$ is a continuous function of a continuous variable $f_c$. It is convenient to express the energy function of a Gibbs distribution as the sums of several terms each associated to cliques of a certain size, that is

$$H(\mathbf{f}) = \sum_{\{i\}\in\mathcal{C}_1} V_1(f_i) + \sum_{\{i,i'\}\in\mathcal{C}_2} V_2(f_i, f_{i'}) + \sum_{\{i,i',i''\}\in\mathcal{C}_3} V_3(f_i, f_{i'}, f_{i''}) + \cdots . \tag{4.12}$$

The above formalization in (4.12) implies a homogeneous Gibbs distribution because the clique potential $V_1, V_2, V_3$ are independent of the locations $i, i'$ and $i''$. For a nonhomogeneous case a cliques potentials should also involve as a parameters sites of the considered cliques as $V_1(i, f_i), V_2(i, i', f_i, f_{i'})$ and so on. An important special case is when only cliques of size up to two are considered (i.e. $V_1$ and $V_2$) due to simple form and low computational cost when it comes to inference i.e. finding the most probable configuration. The well-known MRF models, where energy function is defined in terms of clique potentials of up to two sites, are auto-models [Besag 74]. As a special case of auto-logistic model is an Ising model which we will use in defining our signal model in the subsection 4.2.3.

In 1971 Hammersley and Clifford in their theorem established an equivalence between the MRF and GRF. The local property (Markovianity) which characterizes an MRF and global property of GRF through the energy function $H(\mathbf{f})$ are brought into the equivalence by this theorem. The theorem states that $\mathbf{F}$ is a Markov random field on $\mathcal{S}$ w.r.t. $\mathcal{N}$ if and only if $\mathbf{F}$ is a Gibbs random field on $\mathcal{S}$ w.r.t. $\mathcal{N}$. Many proofs of the theorem exist and can be found in [Besag 74, Moussouris 74, Winkler 12]. Since we will define the MRF model in 4.2.3 starting from the GRF and definition of potential functions, we provide in the following proof that a GRF is an MRF.

Let us denote with $P(\mathbf{f})$ a Gibbs distribution on sites $\mathcal{S}$ w.r.t the neighborhood system $\mathcal{N}$. Consider the conditional probability:

$$P(f_i|\mathbf{f}_{\mathcal{S}-\{i\}}) = \frac{P(f_i, \mathbf{f}_{\mathcal{S}-\{i\}})}{P(\mathbf{f}_{\mathcal{S}-\{i\}})} = \frac{P(\mathbf{f})}{\sum_{f_i' \in \mathcal{L}} P(\mathbf{f}')} \tag{4.13}$$

where $\mathbf{f}' = \{f_1, f_2, ..., f_{i-1}, f_i', f_{i+1}, ..., f_d\}$ is any configuration that agrees with $\mathbf{f}$ at all sites except possible $i$, in other words a $P(f_{\mathcal{S}-\{i\}})$ is obtained by procedure of marginalization by which the random variable $F_i$ has been 'marginalized out' from the joint distribution. Writing out $P(\mathbf{f}) = Z^{-1} \times e^{-\sum_{c \in \mathcal{C}} V_c(\mathbf{f})}$, where we omit $\frac{1}{T}$ for notation simplicity for a moment, gives:

$$P(f_i|\mathbf{f}_{\mathcal{S}-\{i\}}) = \frac{e^{-\sum_{c \in \mathcal{C}} V_c(\mathbf{f})}}{\sum_{f_i' \in \mathcal{L}} e^{-\sum_{c \in \mathcal{C}} V_c(\mathbf{f}')}} \tag{4.14}$$

Then if we divide the set $\mathcal{C}$ into two sets, a set $\mathcal{A}$ consisting of cliques containing site $i$ and a set $\mathcal{B}$ having only cliques not containing $i$, (4.14) can be written as:

$$P(f_i|\mathbf{f}_{\mathcal{S}-\{i\}}) = \frac{[e^{-\sum_{c \in \mathcal{A}} V_c(\mathbf{f})}][e^{-\sum_{c \in \mathcal{B}} V_c(\mathbf{f})}]}{\sum_{f_i' \in \mathcal{L}} \left\{ [e^{-\sum_{c \in \mathcal{A}} V_c(\mathbf{f}')}][e^{-\sum_{c \in \mathcal{B}} V_c(\mathbf{f}')}] \right\}}. \tag{4.15}$$

Since the following equality $V_c(\mathbf{f}) = V_c(\mathbf{f}')$ is valid for any clique $c$ that does not contain site $i$, the $e^{-\sum_{c \in \mathcal{B}} V_c(\mathbf{f})}$ cancels from both the numerator and denominator. Therefore the conditional probability depends only on the potentials of the cliques containing $i$:

$$P(f_i|\mathbf{f}_{\mathcal{S}-\{i\}}) = \frac{e^{-\sum_{c \in \mathcal{A}} V_c(\mathbf{f})}}{\sum_{f_i' \in \mathcal{L}} e^{-\sum_{c \in \mathcal{A}} V_c(\mathbf{f}')}}. \tag{4.16}$$

that is, it depends on labels at $i$'s neighbors. This proves that a Gibbs random field is a Markov random field. One of the very important values of the theorem is that it provides a way of specifying the joint probability distribution. By specifying the clique potential functions $V_c(\mathbf{f})$ with their appropriate form, to resemble the desired system behavior, we can specify the joint probability of the particular system configuration as

$$P(\mathbf{F} = \mathbf{f}) = \frac{1}{Z} e^{-\frac{1}{T} \sum_{c \in \mathcal{C}} V_c(\mathbf{f})}. \tag{4.17}$$

The equation (4.17) characterizes the global context (joint distribution of all pixels in an image) in terms of local spatial characteristics that

are expressed through clique potentials $V_c(\mathbf{f})$. This is the way how an a priori knowledge or preference about interactions between the labeles can be incorporated. In MRF modeling finding the suitable form of a potential functions and determine its parameters for a proper encoding of the system structure is a major topic. In the following section we will present the way how a a priori knowledge about the position of image coefficients under sparsifying transfom can improve signal model and helps in better understanding of signal structure.

## 4.2 Modelling structured sparsity

There are two principal approaches to modeling *structured sparsity* (structure of the sparse image coefficients): (1) in the acquisition stage, through improved design of the sampling trajectories, and (2) in the recovery phase, through an improved regularization of the inverse problem. The first approach is advocated in [Adcock 13, Roman 14, Adcock 14] where efficient multilevel sampling schemes are constructed, showing great potential over the standard sampling strategies. We focus on the second approach – modeling signal structure in the recovery phase.

Recent work has shown benefits of using *wavelet-tree* structure in the MRI recovery [Chen 12, Chen 14]. This approach models the dependencies among wavelet coefficients on a quadtree structure through an additional group sparsity regularization term. For more details about the method see the subsection 3.4.5. Other related approaches employ Hidden Markov Tree (HMT) models [He 09], [Cevher 10b], [Kyrillidis 15]. Less attention has been devoted to modeling within-band (intrascale) dependencies in image recovery from compressive measurements. A representative of this approach is the LaMP (Lattice Matching Pursuit) algorithm [Cevher 09b], where an MRF prior models images that are canonically sparse (in applications such as background subtraction and moving object detection). We give a brief review of the LaMP method in the subsection 3.4.6 where we introduce a related algorithm from [Pižurica 11] that applies an MRF prior to subband data in CS-MRI recovery. Motivated by the encouraging results of [Pižurica 11], we build further on this approach and present a solid motivation, elaborate analysis, and thorough evaluation, while previously only a proof of concept was given. Moreover using a different underlying optimization method and improved MRF modeling, we improve the performance over [Pižurica 11], and we also demonstrate, for the first time, potential benefits over the competing tree-structured approach.

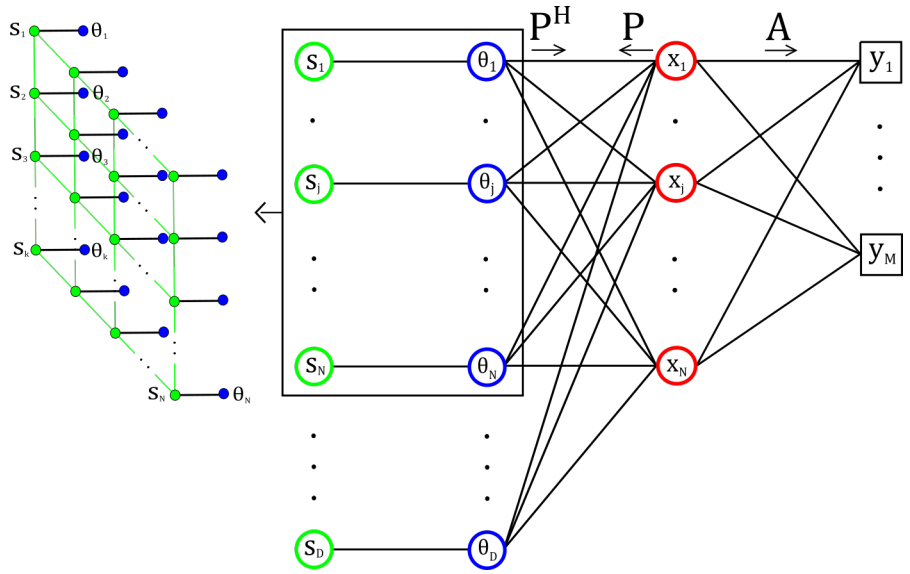



**Figure 4.3:** A graphical representation of variables, operators and their connections in our model. **Left**: Hidden labels $s_i$, attached to each subband coefficient $\theta_i$. Links among neighboring $s_i$ indicate their statistical dependencies, encoded in a MRF. **Right**: A graphical model showing all the involved variables, measurements and operators in our problem.

### 4.2.1 MRF-based structure sparsity model

Let $\mathbf{P}^{D\times N}$ denote some sparsifying transform which yields coefficients $\boldsymbol{\theta} = \mathbf{P}\mathbf{x} = \{\theta_1, ..., \theta_D\}$. The coefficient $\theta_i$ is *significant* if its magnitude is above a certain threshold. We assign a hidden label $s_i \in \{0, 1\}$ to $\theta_i$ to mark its significance: $s_i = 1$ if $\theta_i$ is significant and $s_i = 0$ otherwise. A particular configuration $\mathbf{s} = \{s_1, ..., s_D\}$ is assumed to be a realization of a Markov Random Field $\mathbf{S} = \{S_1, ..., S_D\}$. Fig. 4.3 illustrates this whole setup. Note that all measurements gathered in $\mathbf{y}$ are obtained as linear combinations of all $N$ pixel intensities in $\mathbf{x}$ through the operator $\mathbf{A}$. The sparse coefficients $\boldsymbol{\theta}$ result from applying the analysis operator $\mathbf{P}$ to $\mathbf{x}$. Therefore, each coefficient $\theta_i$ is a linear combination of all pixel values, via $\mathbf{P}$. Conversely, each pixel value $x_i$ is obtained as a linear combination of all coefficients $\theta_i$ through the synthesis operator $\mathbf{P}^H$.

### 4.2.2 Recovery problem with structured sparsity

Let us now instantiate a general recovery problem as (3.23), by replacing the arbitrary regularizer $\phi$ by our structured sparsity model. We use sim-

ilar notation to [Kyrillidis 15]. Given the index set $\mathcal{N} = \{1, 2, 3, ..., D\}$, let $supp(\boldsymbol{\theta}) = \{i \in \mathcal{N} : \theta_i \neq 0\}$ denote the *support* of $\boldsymbol{\theta}$. Further on, for $\mathcal{S} \subseteq \mathcal{N}$, $\boldsymbol{\theta}[\mathcal{S}]$ denotes the elements of $\boldsymbol{\theta}$ indexed by $\mathcal{S}$, and $\bar{\mathcal{S}}$ is the complement of $\mathcal{S}$ with respect to $\mathcal{N}$. Denote the index set corresponding to the support $\mathbf{s}$ by $\Omega_{\mathbf{s}} = \{i \in \mathcal{N} : s_i = 1\}$ and define a model for $\boldsymbol{\theta}$ that conforms to the particular support configuration $\mathbf{s}$ as

$$\mathcal{M}_{\mathbf{s}} = \{\boldsymbol{\theta} \in \mathbb{C}^D : supp(\boldsymbol{\theta}) = \Omega_{\mathbf{s}}\}. \tag{4.18}$$

The objective of our approach is

$$\min_{\mathbf{x} \in \mathbb{C}^N} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \quad \text{subject to} \quad \mathbf{P}\mathbf{x} \in \mathcal{M}_{\hat{\mathbf{s}}} \tag{4.19}$$

where $\hat{\mathbf{s}}$ is the estimate of the most likely spatial support of $\boldsymbol{\theta} = \mathbf{P}\mathbf{x}$. The constraint $\mathbf{P}\mathbf{x} \in \mathcal{M}_{\hat{\mathbf{s}}}$ can be equivalently replaced by $supp(\mathbf{P}\mathbf{x}) = \Omega_{\hat{\mathbf{s}}}$. In solving this problem, we shall involve a simpler one

$$\min_{\boldsymbol{\gamma} \in \mathbb{C}^D} \|\boldsymbol{\gamma} - \boldsymbol{\theta}\|_2^2 \quad \text{subject to} \quad \boldsymbol{\gamma} \in \mathcal{M}_{\hat{\mathbf{s}}} \tag{4.20}$$

for which the solution is $\hat{\boldsymbol{\gamma}}_H[\Omega_{\hat{\mathbf{s}}}] = \boldsymbol{\theta}[\Omega_{\hat{\mathbf{s}}}]$ and $\hat{\boldsymbol{\gamma}}_H[\bar{\Omega}_{\hat{\mathbf{s}}}] = 0$. Since $s_i \in \{0, 1\}$, this solution can be written as the *Hadamard product* $\hat{\boldsymbol{\gamma}}_H[\Omega_{\hat{\mathbf{s}}}] = \boldsymbol{\theta} \circ \hat{\mathbf{s}}$.

We search for the most likely support $\hat{\mathbf{s}}$ by applying the maximum a posteriori probability (MAP) criterion:

$$\hat{\mathbf{s}} = \underset{\mathbf{s}}{\operatorname{argmax}}\, P_{\mathbf{S}|\boldsymbol{\theta}}(\mathbf{s} \mid \boldsymbol{\theta}) = \underset{\mathbf{s}}{\operatorname{argmax}}\, p_{\boldsymbol{\theta}|\mathbf{S}}(\boldsymbol{\theta} \mid \mathbf{s}) P_{\mathbf{S}}(\mathbf{s}) \tag{4.21}$$

In practice, we shall re-estimate $\hat{\mathbf{s}}$ in each iteration of the complete recovery algorithm, starting from the current (temporary) estimate of the coefficient vector $\boldsymbol{\theta}$.

### 4.2.3 MRF prior

According to Hammersley-Clifford theorem introduced in subsection 4.1.1 the global probability $P_{\mathbf{S}}(\mathbf{s})$ of a MRF is a Gibbs distribution [Li 09], [Pižurica 02]

$$P_{\mathbf{S}}(\mathbf{s}) = \frac{1}{Z} e^{-H(\mathbf{s})/T} \tag{4.22}$$

where the energy $H(\mathbf{s})$ is a sum of clique potentials over all possible cliques: $H(\mathbf{s}) = \sum_{c \in \mathcal{C}} V_c(\mathbf{s})$. This allows for an efficient factorization

of the Gibbs distribution using local interactions modeled by potential functions. We use the Ising model as in [Pižurica 11], which is characterized by encoding the local interactions by pair-site clique potentials. The Ising model, as a simplified version of auto-logistic model, has the energy function of the following form

$$H(\mathbf{s}) = \sum_i V_1(s_i) + \sum_{\langle i,j\rangle \in \mathcal{C}} V_2(s_i, s_j) \tag{4.23}$$

with the single and pairwise potentials defined as

$$V_1(s) = \begin{cases} \alpha & s = 0 \\ -\alpha & s = 1 \end{cases}, \quad V_2(s,t) = \begin{cases} -\beta & s = t \\ \beta & s \neq t \end{cases} \tag{4.24}$$

Unlike in [Pižurica 11], we allow different a priori probabilities $\alpha \neq 0$, so that we can enforce the sparsity of the supports. The strength of the spatial clustering is controlled by the parameter $\beta > 0$.

### 4.2.4 Conditional model

We adopt the conditional model $p_{\mathbf{\Theta}|\mathbf{S}}(\boldsymbol{\theta}|\mathbf{s})$ of [Pižurica 11, Pižurica 02]. With the common conditional independence assumption, we have $p_{\mathbf{\Theta}|\mathbf{S}}(\boldsymbol{\theta}|\mathbf{s}) = \prod_i p_{\Theta_i|S_i}(\theta_i|s_i)$. The observed coefficients are typically noisy versions of the ideal ones: $\theta = u + n$, where $n$ denotes the noise component. We select the prior $p_U(u)$ as the generalized Laplacian and we estimate its parameters from the noisy coefficient histogram, knowing the noise standard deviation $\sigma$ [Pižurica 02, Simoncelli 97]. In practice, $\sigma$ is reliably estimated from the empty area on the borders of the MR image and rescaled appropriately in each subband. Let $T_h$ denote the significance threshold for the ideal noise-free coefficients ($u$ is significant if $|u| \geq T_h$). We relate this threshold to the noise level, but in a conservative manner, such that $T_h$ is only a fraction of $\sigma$ (in practice 10%). The conditional densities $p_{U|S}(u|0)$ and $p_{U|S}(u|1)$ are then obtained by rescaling the central part ($|u| < T_h$) and the tails ($|u| \geq T_h$) of $p_U(u)$, respectively, so that they both integrate to 1. The conditional densities of the noisy coefficients $p_{\Theta|S}(\theta|s)$ are obtained from the corresponding $p_{U|S}(u|s)$. For the additive noise model $\theta = u + n$ with $n \sim N(0, \sigma)$, $p_{\Theta|S}(\theta|s)$ is simply the convolution of $p_{U|S}(u|s)$ with $N(0, \sigma)$. Fig. 4.4 illustrates the adopted conditional model and the above described procedure.

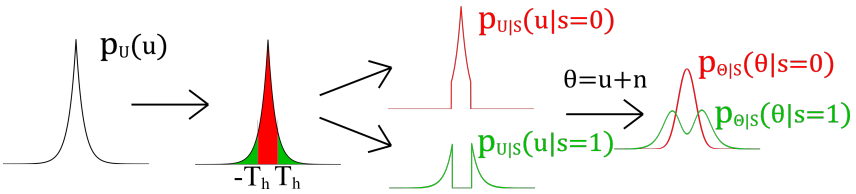



**Figure 4.4:** The adopted conditional model from [Pižurica 11], [Pižurica 02]. Note that $p_U(u)$ is obtained from the noisy histogram. $T_h$ is the only parameter.

### 4.2.5 Inference algorithm

Various inference algorithms can be employed to find the MAP estimate in (4.21), e.g., Iterative Conditional Modes (ICM) [Besag 86], Graph Cuts [Kolmogorov 04], loopy belief propabation (LBP) [Murphy 99], and Markov Chain Monte Carlo (MCMC) samplers, such as Metropolis and Gibbs sampler [Li 09]. We used the Metropolis sampler due to its flexibility and efficiency in this application. The Metropolis sampler starts from some initial configuration and in each step it switches a randomly chosen label $s_i$ in the current mask $\mathbf{s}$ to produce the so-called "candidate" mask $\mathbf{s}^C$. The candidate gets accepted or not based on the change in the posterior probability $P_{\mathbf{S}|\boldsymbol{\Theta}}(\mathbf{s}^C|\boldsymbol{\theta})/P_{\mathbf{S}|\boldsymbol{\Theta}}(\mathbf{s}|\boldsymbol{\theta})$, which effectively reduces to

$$r = \left(\frac{p_{\theta_i|S_i}(\theta_i \mid 1)}{p_{\theta_i|S_i}(\theta_i \mid 0)}\right)^{\lambda} \exp\left\{2\alpha + 2\beta \sum_{j\in\mathcal{N}_i}(2s_j - 1)\right\} \tag{4.25}$$

when $s_i^C = 1$ and to $1/r$ when $s_i^C = 0$. Practically, the change is accepted if $r$ exceeds a randomly generated number drawn from a uniform distribution on $[0, 1]$. Parameter $\lambda > 0$ effectively simulates sampling at different temperatures; for details see [Pižurica 02]. This inference algorithm is in fact a step of the simulated annealing algorithm from [Kirkpatrick 84] for a particular temperature — one could apply simulated annealing by changing gradually $\lambda$ although we didn't do it in our experiments.

## 4.3 LaSAL

Our optimization problem from (4.19) is equivalent to (3.23) under suitably defined regularization function $\phi$. We follow the same steps for solving (3.23), as explained in derivation of CSALSA method [Afonso 11] in 3.4.3, and we incorporate the particular $\phi$ that follows from our structured sparsity model described in the previous chapter. To this end, let

$E(\epsilon, \mathbf{A}, \mathbf{y}) = \{\mathbf{x} \in \mathbb{C}^N : \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon\}$ denote the feasible set for $\mathbf{x}$. By introducing an indicator function

$$\iota_{\mathcal{Q}}(\mathbf{q}) = \begin{cases} 0, & \mathbf{q} \in \mathcal{Q} \\ +\infty, & \text{otherwise} \end{cases} \tag{4.26}$$

the problem in (3.23) can be written as follows:

$$\min_{\mathbf{x} \in \mathbb{C}^N} \phi(\mathbf{P}\mathbf{x}) + \boldsymbol{\iota}_{E(\epsilon, \mathbf{I}, \mathbf{y})}(\mathbf{A}\mathbf{x}). \tag{4.27}$$

In the subsection 3.4.3 it has been shown that this problem is efficiently solved by a special type of the alternating direction method of multipliers (ADMM) [Afonso 11]. The key step is variable splitting technique introduced in subsection 3.4.2, which allows solving the composite problem as a sequence of minimizations over the separate components. In particular, for the problem in (4.27), two splitting variables are introduced $\mathbf{w} = \mathbf{x}$ and $\mathbf{v} = \mathbf{A}\mathbf{x}$, to split the original problem into separate mininimizations over each of the two terms. Together with a "binding" term that connects these two separate minimizations, we obtain the following three sub-problems:

$$\begin{aligned} \mathbf{x}^{\{k+1\}} &= \underset{\mathbf{x} \in \mathbb{C}^N}{\operatorname{argmin}} \left\{ \|\mathbf{A}\mathbf{x} - \mathbf{u}'\|_2^2 + \mu \|\mathbf{x} - \mathbf{u}''\|_2^2 \right\} \\ \mathbf{v}^{\{k+1\}} &= \underset{\mathbf{v} \in \mathbb{C}^M}{\operatorname{argmin}} \left\{ \frac{\boldsymbol{\iota}_{E(\epsilon, \mathbf{I}, \mathbf{y})}(\mathbf{v})}{\mu} + \frac{1}{2} \|\mathbf{v}' - \mathbf{v}\|_2^2 \right\} \\ \mathbf{w}^{\{k+1\}} &= \underset{\mathbf{w} \in \mathbb{C}^N}{\operatorname{argmin}} \left\{ \phi(\mathbf{P}\mathbf{w}) + \frac{\mu}{2} \|\mathbf{w}' - \mathbf{w}\|_2^2 \right\} \end{aligned} \tag{4.28}$$

where $\mathbf{v}' = \mathbf{A}\mathbf{x}^{\{k+1\}} - \mathbf{b}^{\{k\}}$, $\mathbf{w}' = \mathbf{x}^{\{k+1\}} - \mathbf{c}^{\{k\}}$, $\mathbf{u}' = \mathbf{v}^{\{k\}} + \mathbf{b}^{\{k\}}$, $\mathbf{u}'' = \mathbf{w}^{\{k\}} + \mathbf{c}^{\{k\}}$, and $\mathbf{b}, \mathbf{c}$ are auxiliary variables.

The first sub-problem $\mathbf{x}^{\{k+1\}}$ is solved by the Gauss-Seidel method leading to a simple update equation. The second sub-problem $\mathbf{v}^{\{k+1\}}$ obviously does not depend on $\mu$ (because the indicator function defined in (4.26) takes only the values 0 or $+\infty$) and is simply the orthogonal projection of $\mathbf{v}$ on the closed $\epsilon$-radius ball centered at $\mathbf{y}$ [Afonso 11]:

$$\boldsymbol{\Psi}_{\iota_{E(\epsilon, \mathbf{I}, \mathbf{y})}}(\mathbf{v}) = \mathbf{y} + \begin{cases} \epsilon \frac{\mathbf{v} - \mathbf{y}}{\|\mathbf{v} - \mathbf{y}\|_2}, & \text{if } \|\mathbf{v} - \mathbf{y}\|_2 > \epsilon \\ \mathbf{v} - \mathbf{y}, & \text{if } \|\mathbf{v} - \mathbf{y}\|_2 \leq \epsilon \end{cases} \tag{4.29}$$

The third sub-problem $\mathbf{w}^{\{k+1\}}$ has been typically solved by defining $\phi$ as the $\ell_1$-norm. We define instead the regularization function $\phi(\boldsymbol{\theta})$ as a $\delta$-loss function, prohibiting all realizations $\boldsymbol{\theta}$ that do not

**Algorithm 11** LaSAL

**Input:** $k = 0, \mu > 0, \mathbf{v}^{\{0\}}, \mathbf{w}^{\{0\}}, \mathbf{b}^{\{0\}}, \mathbf{c}^{\{0\}}$

**repeat**
  $\mathbf{r}^{\{k\}} = \mu(\mathbf{w}^{\{k\}} + \mathbf{c}^{\{k\}}) + \mathbf{A}^H(\mathbf{v}^{\{k\}} + \mathbf{b}^{\{k\}})$
  $\mathbf{x}^{\{k+1\}} = (\mu\mathbf{I} + \mathbf{A}^H\mathbf{A})^{-1}\mathbf{r}^{\{k\}}$
  $\mathbf{v}^{\{k+1\}} = \mathbf{\Psi}_{\iota_{E(\epsilon,\mathbf{I},\mathbf{y})}}(\mathbf{A}\mathbf{x}^{\{k+1\}} - \mathbf{b}^{\{k\}})$
  $\boldsymbol{\theta}' = \mathbf{P}(\mathbf{x}^{\{k+1\}} - \mathbf{c}^{\{k\}})$
  $\hat{\mathbf{s}} \leftarrow \mathit{MAP\text{-}support}\{\boldsymbol{\theta}'\}$
  $\mathbf{w}^{\{k+1\}} = \mathbf{P}^H(\boldsymbol{\theta}' \circ \hat{\mathbf{s}})$
  $\mathbf{b}^{\{k+1\}} = \mathbf{b}^{\{k\}} - (\mathbf{A}\mathbf{x}^{\{k+1\}} - \mathbf{v}^{\{k+1\}})$
  $\mathbf{c}^{\{k+1\}} = \mathbf{c}^{\{k\}} - (\mathbf{x}^{\{k+1\}} - \mathbf{w}^{\{k+1\}})$
  $k = k + 1$
**until** some stopping criterion is satisfied

conform to the estimated support $\hat{\mathbf{s}}$. With the model $\mathcal{M}_{\hat{\mathbf{s}}}$ from (4.18), we define formally

$$\phi(\boldsymbol{\theta}) = \begin{cases} 0, & \text{if } \boldsymbol{\theta} \in \mathcal{M}_{\hat{\mathbf{s}}} \\ \infty, & \text{if } \boldsymbol{\theta} \notin \mathcal{M}_{\hat{\mathbf{s}}} \end{cases} \tag{4.30}$$

Substituting $\boldsymbol{\theta} = \mathbf{P}\mathbf{w}$, the third sub-problem in (4.28), following the transformation procedure given in [Liu 16], becomes

$$\boldsymbol{\theta}^{\{k+1\}} = \underset{\boldsymbol{\theta} \in \mathit{Range}(\mathbf{P})}{\operatorname{argmin}} \left\{ \phi(\boldsymbol{\theta}) + \frac{\mu}{2}\|\boldsymbol{\theta}' - \boldsymbol{\theta}\|_2^2 \right\} \tag{4.31}$$

This problem has the same solution as its equivalent constrained formulation from (4.20), and thus $\boldsymbol{\theta}^{\{k+1\}} = \boldsymbol{\theta}' \circ \hat{\mathbf{s}}$ where

$$[\boldsymbol{\theta}' \circ \hat{\mathbf{s}}]_i = \begin{cases} \theta_i, & \text{if } \hat{s}_i = 1 \\ 0, & \text{if } \hat{s}_i = 0 \end{cases} \tag{4.32}$$

This completes the specification of our algorithm, named by analogy with the related methods as LaSAL, from Lattice Split Augmented Lagrangian. Its pseudo-code is listed in Algorithm 11. The step $\hat{\mathbf{s}} \leftarrow \mathit{MAP\text{-}support}\{\boldsymbol{\theta}'\}$ denotes the support estimation using the MAP criterion in (4.21). The parameter $0 < \mu \leq 1$, which controls the level of regularization, can be safely set to 1 as it was also done in [Afonso 11], without a significant performance loss. We still decided to keep $\mu$ as a parameter in the algorithm, because we observed that allowing values of $\mu \leq 1$ can yield a slightly higher peak signal to noise ratio (PSNR) in the

reconstructions (up to 0.5 dB). Furthermore, observe that the update of the auxiliary variable $\mathbf{c}$ (line 9) is performed in the image-domain (while the equivalent step of the analysis-formulation of the related C-SALSA is applied in the transform domain, and with $\mathbf{P}^H$ applied to the first sum in $\mathbf{r}^{\{k\}}$). This is because we use $\mathbf{w} = \mathbf{x}$ in variable splitting instead of $\mathbf{w} = \mathbf{P}\mathbf{x}$ which leads to this type of update. Finally, as the stopping criterion, we use in practice a fixed number of iterations (typically 50), because in all the simulations the differences in the resulting reconstruction error become practically negligible after this many iterations.

## 4.4 LaSAL2

We extend now the objective function in (4.27) with another regularization term: TV norm $\|\mathbf{x}\|_{\mathrm{TV}}$. The resulting objective function is an instance of the general form

$$\min_{\mathbf{x}\in\mathbb{C}^N} \sum_{j=1}^{J} g_j(\mathbf{H}^{(j)}\mathbf{x}) \tag{4.33}$$

with $J = 3$, $\mathbf{H}^{(1)} = \mathbf{P}$, $\mathbf{H}^{(2)} = \mathbf{I}$, $\mathbf{H}^{(3)} = \mathbf{A}$, $g_1(\mathbf{u}) = \phi(\mathbf{u})$, $g_2(\mathbf{u}) = \|\mathbf{u}\|_{\mathrm{TV}}$ and $g_3(\mathbf{u}) = \iota_{E(\epsilon,\mathbf{I},\mathbf{y})}(\mathbf{u})$. A detailed explanation and a compact pseudo-code for solving (4.33) in general case is reviewed from [Afonso 11] in subsection 3.4.3. Here, we simply extend LaSAL from Algorithm 11 with an additional step that concerns with the TV regularization. This also requires introducing an additional auxiliary variable ($\mathbf{d}^{\{k\}}$, next to $\mathbf{b}^{\{k\}}$ and $\mathbf{c}^{\{k\}}$ in Algorithm 11). By applying variable splitting approach [Afonso 09] explained in 3.4.2, the minimization sub-problem corresponding to the TV regularization can be written as

$$\begin{aligned} \mathbf{z}^{\{k+1\}} &= \underset{\mathbf{g}\in\mathbb{C}^N}{\operatorname{argmin}} \left\{ \|\mathbf{z}\|_{\mathrm{TV}} + \frac{\mu_1+\mu_2}{2}\|\mathbf{z}' - \mathbf{z}\|_2^2 \right\} \\ &= \mathbf{\Psi}_{\mathrm{TV}}(\mathbf{z}'; \mu_1+\mu_2) \end{aligned} \tag{4.34}$$

where $\mathbf{z}' = \frac{1}{(\mu_1+\mu_2)}\big(\mu_1(\mathbf{x}^{\{k+1\}} - \mathbf{c}^{\{k\}}) + \mu_2(\mathbf{w}^{\{k\}} + \mathbf{d}^{\{k\}})\big)$ is a linear combination of the current solution $\mathbf{x}^{\{k+1\}}$ and the regularized solution $\mathbf{w}^{\{k\}}$ from the previous iteration, with parameters where $\mu_1, \mu_2 > 0$. For computing $\mathbf{\Psi}_{\mathrm{TV}}(\mathbf{u}; \mu)$ we used 5 iterations of Chambolle's algorithm [Chambolle 04] (more iterations only increased the computational cost with negligible improvement in PNSR). A pseudo-code of the resulting

**Algorithm 12** LaSAL2

**Input:** $k = 0, \mu_1, \mu_2 > 0, \mathbf{v}^{\{0\}}, \mathbf{w}^{\{0\}}, \mathbf{z}^{\{0\}}, \mathbf{b}^{\{0\}}, \mathbf{c}^{\{0\}}, \mathbf{d}^{\{0\}},$

**repeat**
- $\mathbf{r}^{\{k\}} = \mu_1(\mathbf{z}^{\{k\}} + \mathbf{c}^{\{k\}}) + \mathbf{A}^H(\mathbf{v}^{\{k\}} + \mathbf{b}^{\{k\}})$
- $\mathbf{x}^{\{k+1\}} = (\mu_1\mathbf{I} + \mathbf{A}^H\mathbf{A})^{-1}\mathbf{r}^{\{k\}}$
- $\mathbf{v}^{\{k+1\}} = \mathbf{\Psi}_{\iota_{E(\epsilon,\mathbf{I},\mathbf{y})}}(\mathbf{A}\mathbf{x}^{\{k+1\}} - \mathbf{b}^{\{k\}})$
- $\mathbf{z}' = \frac{1}{(\mu_1+\mu_2)}\big(\mu_1(\mathbf{x}^{\{k+1\}} - \mathbf{c}^{\{k\}}) + \mu_2(\mathbf{w}^{\{k\}} + \mathbf{d}^{\{k\}})\big)$
- $\mathbf{z}^{\{k+1\}} = \mathbf{\Psi}_{\mathrm{TV}}(\mathbf{z}'; \mu_1 + \mu_2)$
- $\boldsymbol{\theta}' = \mathbf{P}(\mathbf{z}^{\{k+1\}} - \mathbf{d}^{\{k\}})$
- $\hat{\mathbf{s}} \leftarrow \mathit{MAP\text{-}support}\{\boldsymbol{\theta}'\}$
- $\mathbf{w}^{\{k+1\}} = \mathbf{P}^H(\boldsymbol{\theta}' \circ \hat{\mathbf{s}})$
- $\mathbf{b}^{\{k+1\}} = \mathbf{b}^{\{k\}} - (\mathbf{A}\mathbf{x}^{\{k+1\}} - \mathbf{v}^{\{k+1\}})$
- $\mathbf{d}^{\{k+1\}} = \mathbf{d}^{\{k\}} - (\mathbf{z}^{\{k+1\}} - \mathbf{w}^{\{k+1\}})$
- $\mathbf{c}^{\{k+1\}} = \mathbf{c}^{\{k\}} - (\mathbf{x}^{\{k+1\}} - \mathbf{z}^{\{k+1\}})$
- $k = k + 1$

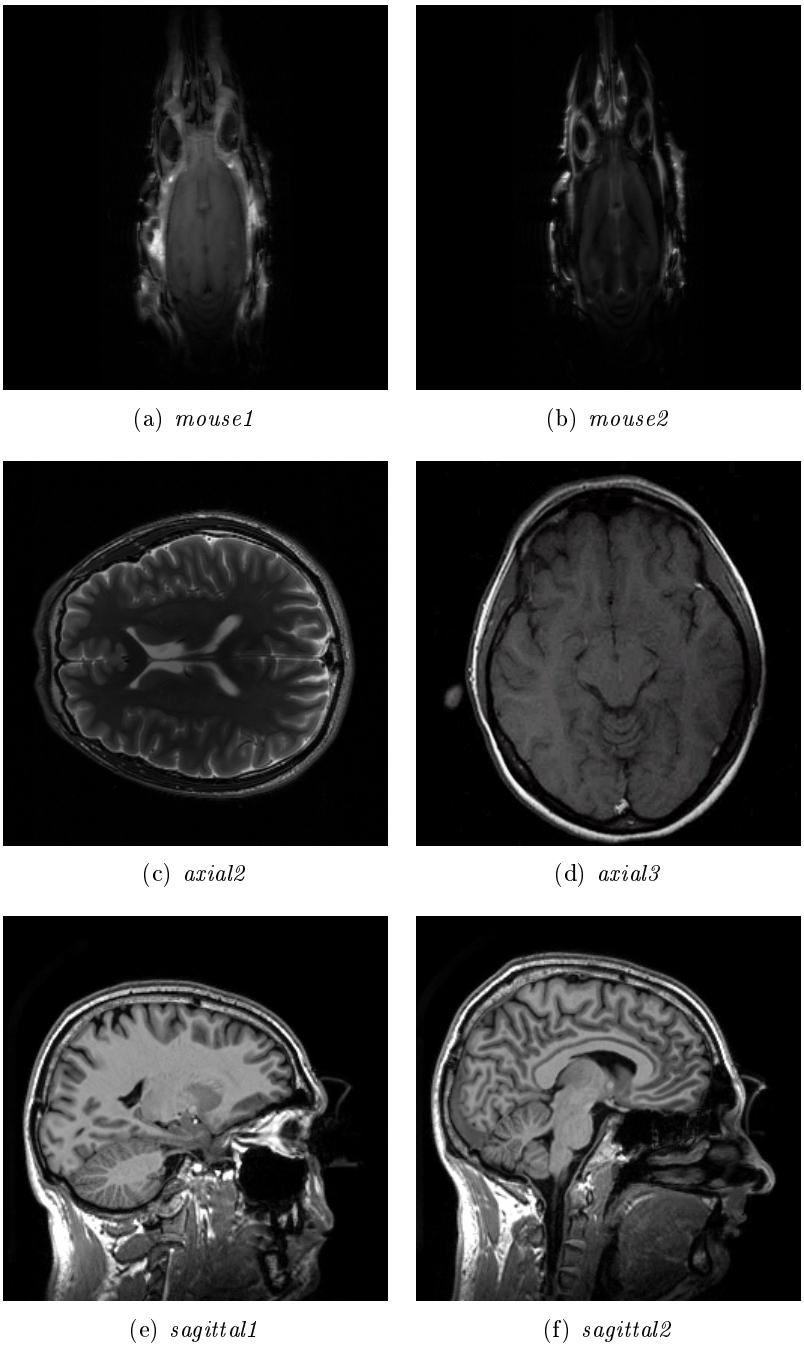
**until** some stopping criterion is satisfied

method that we named LaSAL2 is given in Algorithm 12. The source codes of both LaSAL and LaSAL2 algorithms are available at `https://telin.ugent.be/~sanja/MRIreconstruction/LaSAL`.

## 4.5 Experimental evaluation of LaSAL and LaSAL2

As reference methods, we use C-SALSA [Afonso 11], the augmented Lagrangian method (Split-Bregman) SB [Aelterman 11] and LaSB [Pižurica 11], all implemented with the same non-decimated shearlet transform introduced in 3.3.4. We also provide comparison with WaTMRI [Chen 12, Chen 14] and FCSA [Huang 11b] (see subsections 3.4.5 and 3.4.4) and FCSANL [Huang 12] using the original implementations of the authors (`http://ranger.uta.edu/~huang/index.html`). Comparison with pFISTA [Liu 16], dictionary learning approach DLMRI [Ravishankar 11] and a patch-based method PANO [Qu 14] is provided on images for which these methods were optimized. We use a data set that comprises 248 T1 MRI *brain slices* acquired on a Cartesian grid at Ghent University Hospital (UZ Gent)[1], also used in [Pižurica 11, Aelter-

[1] Data acquired thanks to Prof. Dr. Karel Deblaere at the Radiology Department of UZ Gent.



(a) *mouse1* (b) *mouse2*

(c) *axial2* (d) *axial3*

(e) *sagittal1* (f) *sagittal2*

**Figure 4.5:** The top images *mouse1* and *mouse2* are two slices from images of *mouse brain*. The middle two images *axial2* and *axial3* are from [Qu 14] and [Ravishankar 11] resp. The bottom two images *sagittal1* and *sagittal2* are from *brain slices* dataset. All images are 256×256, except *axial3*, which is 512×512.

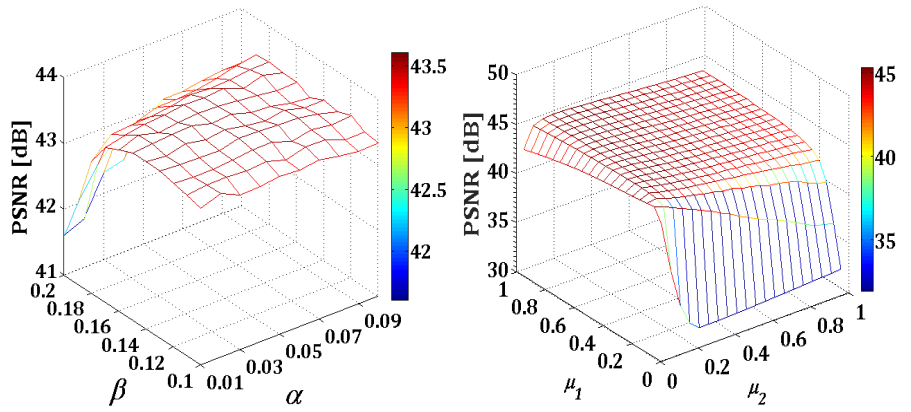



**Figure 4.6:** Examples illustrating grid search results for the MRF parameters $(\alpha, \beta)$ (left) and regularization parameters $(\mu_1,\mu_2)$ (right).

man 11]. All the test images in this data set have resolution 256×256. In our evaluation we also use two images of *mouse brain* with different modalities (T1 and T2) acquired on a Cartesian grid at Bio-Imaging Lab at the University of Antwerp (see Fig. 4.5). As an evaluation criterion, we use the peak signal to noise ratio (PSNR) computed on the magnitude image. For the data acquired directly in $k$-space (no reference image available), we compute the structural similarity index (SSIM) [Wang 04] of the reconstructions from partial data relative to the reconstruction from all the available measurements. SSIM is a perception-based model that considers image degradation as perceived change in structural information which reflects that the pixels have strong inter-dependencies especially when they are spatially close. We show results through the thesis mostly for the reconstruction of image magnitude if not otherwise stated. For images that have important information contained in phase, reconstruction is conducted by independent application of regularization on the real and imaginary part of the image per each algorithm iteration.

### 4.5.1 Parameter selection

As a sparsifying transform for our proposed methods, we use a non-decimated shearlet transform (see subsection 3.3.4). We decide to use the implementation of shearlet transform from [Goossens 09] due to flexibility in the selection of a number of scales and orientations per scale and low computation cost. For the transformation parameter we select 3 scales and by default 16, 8, and 4 orientations per scale, respectively.

The parameters $(\alpha, \beta)$ of the MRF model are optimized by grid search. Although the optimal values may slightly differ depending on the particular image, sampling rate, and sampling trajectory, we observed a stable performance in a relatively wide range of the parameter values, as illustrated in the example from Fig. 4.6 (left). This diagram corresponds to LaSAL, applied on the test image *sagittal1*, with 48% of samples on a radial trajectory. Similar diagrams were obtained with other images and other sampling trajectories. We concluded that the same parameter values can be safely used for a wide range of sampling rates and different images. The recommended values are $\alpha$=0.01; $\beta$=0.16, with $\lambda = 0.2$ in (4.25). All the image reconstruction results reported in this paper were produced with these values. We did observe that somewhat more stable performance is in some cases reached with slightly different parameter values and at the price of slightly reduced PSNR, but these differences are not so significant in our experience.

We also optimize the parameter $\mu$ of LaSAL and $(\mu_1, \mu_2)$ of LaSAL2 by grid search. It is important to note that all these parameters can be simply set to 1, without sacrificing significantly the reconstruction performance. This is evident from the grid search diagram in Fig. 4.6 (right) and agrees also with the general theory in [Afonso 11]. Still, we observed that somewhat better reconstruction performance may be reached in practice by allowing other values of these parameters, so we opted to keep the possibility for their experimental optimization. In particular, we recommend $\mu = 0.04$ for LaSAL, and $\mu_1 = 0.11$, $\mu_2 = 0.01$ for LaSAL2 which are used for all experiments in the paper.

### 4.5.2 Benefit from the MRF model

We first explore how the incorporated MRF-based spatial context model influences the reconstruction performances. This can be directly observed by excluding the MRF-modelling part of LaSAL (lines 5–7 in Algorithm 11) and replacing it simply by soft-thresholding in the shearlet domain $\mathbf{w}^{\{k+1\}} = \mathbf{P}^H(\mathbf{\Psi}_{\ell_1}(\mathbf{P}(\mathbf{x}^{\{k+1\}} - \mathbf{c}^{\{k\}}; \mu)))$, which reduces our method to the corresponding version of C-SALSA [Afonso 11].

The results in Fig. 4.7 demonstrate a clear improvement due to the MRF model. LaSB and SB share the same optimization algorithm, while LaSB is enriched with an MRF model. Similarly, C-SALSA and LaSAL share the same optimization method, extended with an MRF prior in LaSAL. We observe that in the same way as LaSB improves over SB, our new algorithm LaSAL improves over C-SALSA consistently for

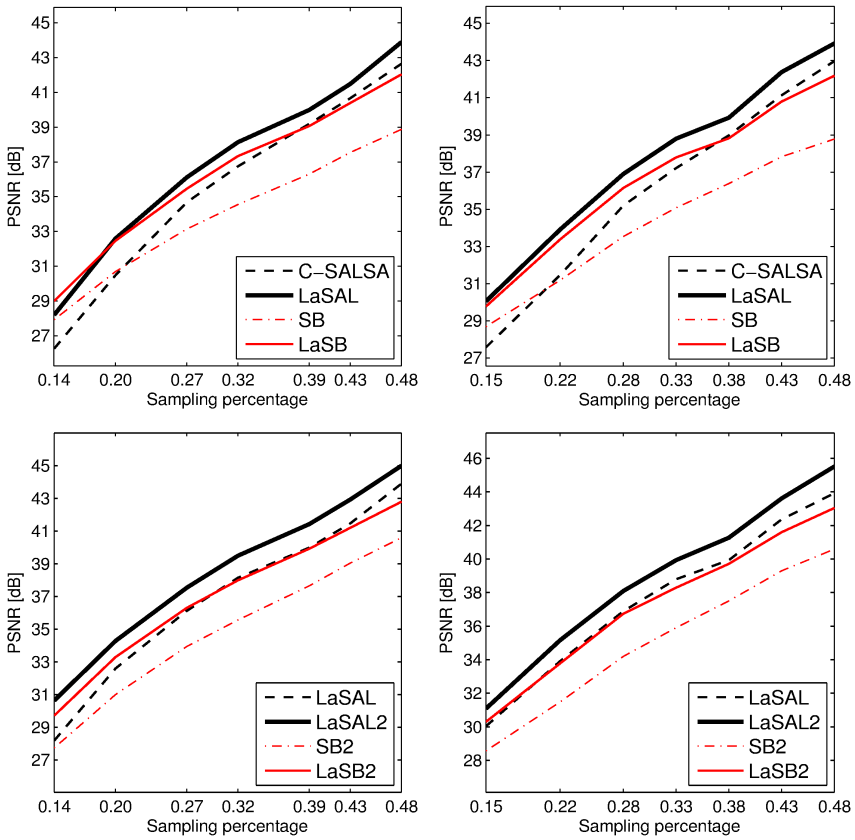



**Figure 4.7:** Reconstruction results for *sagittal1* with different sampling rates using radial (left) and Fibonacci spiral (right) trajectories.

all sampling rates. Moreover, LaSAL yields consistent improvement over LaSB, except at very low sampling rates. With the spiral trajectory, for sampling rates above 0.3, this improvement in PSNR is more than 1 dB and above 1.7 dB for the sampling rate around 0.5. Similar behavior, with slightly smaller differences, is observed in the case of radial trajectory.

Fig. 4.7 (bottom row) also shows the advantage of the compound prior: LaSAL2 indeed improves over LaSAL. The results also demonstrate improvement over the reference methods SB and LaSB implemented with compound priors, denoted for consistency as SB2 and LaSB2. The improvement of LaSAL2 over these methods is consistent at all sampling rates and for both sampling trajectories. The difference

in PSNR relative to both LaSAL and LaSB2 ranges from 1 dB to more than 2 dB, while the improvement over SB2 is 3 to 5 dB.

### 4.5.3 Comparison with other methods

The reference methods FCSA [Huang 11b], FCSANL [Huang 12] and WaTMRI [Chen 14] employ a compound regularization: TV and $\ell_1$ (FCSA and WaTMRI) or non-local TV and $\ell_1$ (FCSANL). WaTMRI employs next to it a tree-structured sparsity model. We adopt the experimental setup of [Chen 14,Huang 11b,Huang 12] using random sampling matrices with variable density[2]. Seven sampling rates (14%, 20%, 25%, 32%, 38%, 42% and 50%) are used, and for each of them ten sampling matrices are randomly generated and the average PSNR over the ten corresponding reconstructions is recorded. Since we tested the reference methods on different images, from those used in [Huang 11b], [Huang 12] [Chen 14], we optimize their parameters accordingly using the grid search approach.

Fig. 4.8 (top left) shows the result for *sagittal1* from Fig. 5.3. Obviously, the proposed LaSAL and LaSAL2 algorithms yield consistent improvement over all three reference methods FCSA, FCSANL and WaTMRI at all sampling rates. This improvement is in the range of 1.4 – 3 dB for LaSAL and in the range of 2.3 – 4.1 dB for LaSAL2. Similar conclusions hold for *sagittal2* (Fig. 4.8, top right): the improvement for LaSAL is now in the range of 0.5 – 2.6 dB and for LaSAL2 the same as on *sagittal1*. Two other diagrams in Fig. 4.8 show the PSNR results for the T1 and T2 *mouse brain* images. LaSAL2 yields again superior performance compared to all other tested methods. Among the three reference methods, FCSANL is now the best performing. LaSAL2 improves over this method at all sampling rates in the range of 0.6 dB to 2.7 dB for *mouse1* and in the range 2.2 – 3 dB for *mouse2*. Visual comparison of selected methods for different sampling rates are shown in Fig. 4.9 and Fig. 4.10. From reconstructed images visually is very hard or even impossible to see the differences, hence we provide a visualization of difference image with respect to the ground truth image from Fig. 4.5. In the difference images we see that proposed methods LaSAL and LaSAL2 have smaller intensities of errors than WaTMRI in the region of interest, e.g, where the signal is present.

We also evaluate on the complete dataset of 248 MRI *brain slices*. Fig. 4.11 shows the mean PSNR per iteration across all the 248 im-

[2] `http://ranger.uta.edu/~huang/index.html`

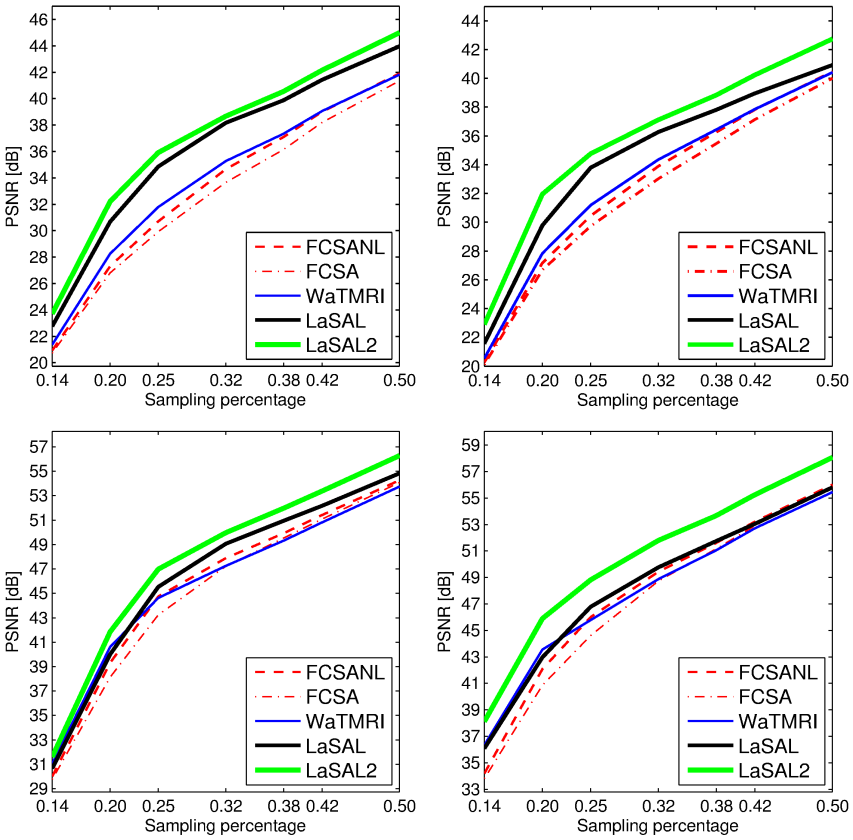



**Figure 4.8:** Results with random sampling and different sampling rates on images *sagittal1* and *sagittal2* (top); *mouse1* and *mouse2* (bottom).

ages for different algorithms. LaSAL2 yields considerably higher PSNR than the reference methods: the improvement is more than 3.5 dB. Even LaSAL, which employs no TV regularization, improves over FCSANL and WaTMRI for about 2.4 dB and reaches its highest PSNR in fewer iterations than the reference methods. In the same figure, we show the resulting distribution of the PSNR values per iteration. It can be seen that after 5 iterations LaSAL reaches a huge improvement in PSNR over all the reference methods, while LaSAL2 outperforms LaSAL after 30 iterations.

This huge improvement in PSNR comes at the price of increased processing time. The computation times reported below were obtained on Intel© Core™ i7 processor (2.4 GHz, 8GB RAM). For LaSAL and

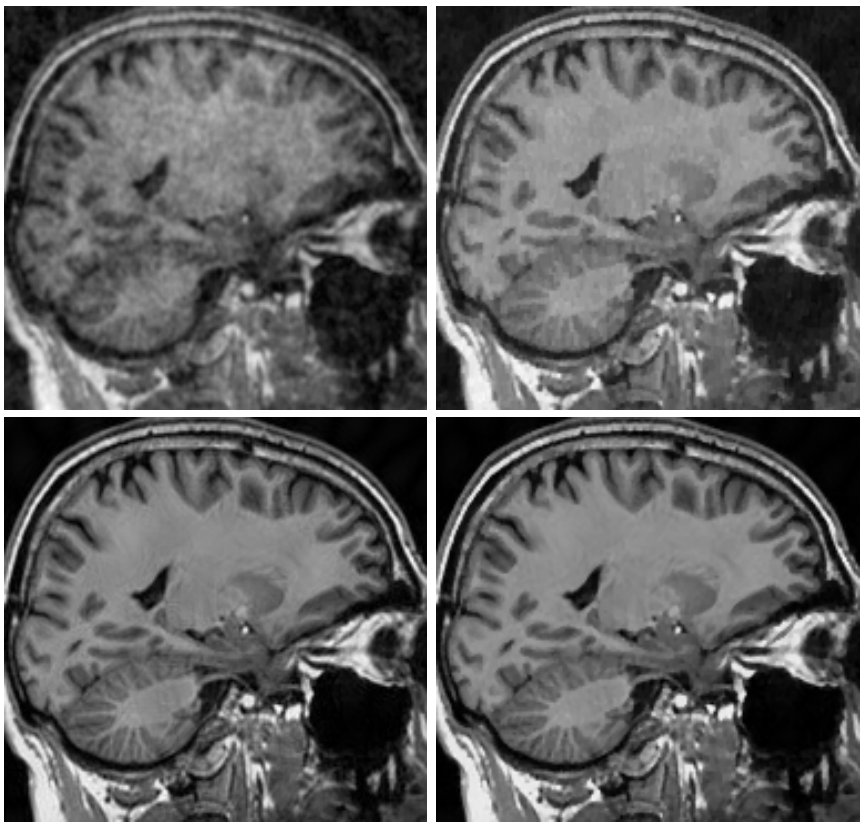


**Figure 4.9:** Reconstructed *sagittal1* image from 20% of random measurements. **First row**: zero-fill (19.87 dB) and WaTMRI (28.78 dB), **Second row**: LaSAL (31.06 dB) and LaSAL2 (33.43 dB).

LaSAL2, with a non-optimized Matlab implementation and (16, 8, 4) shearlet bands per scale, the processing time for a $256 \times 256$ image is 1.50 s per iteration, out of which 1.41 s goes on the support estimation, resulting in about 75 s for 50 iterations. For comparison, the fastest reference methods in 50 iterations require: FCSANL – 0.7 s, FCSA – 0.6 s, WaTMRI – 0.8 s, and C-SALSA – 10.9 s. There is much room for improving the computation time of our method by improving the efficiency of the support configuration, e.g. by considering alternatives to Metropolis sampling, such as iterated conditional modes (ICM) [Besag 86] or belief propagation.

For comparisons with pFISTA [Liu 16], we use their data – *axial1* in Fig. 5.3 and the original code provided by the authors. Fig. 4.12

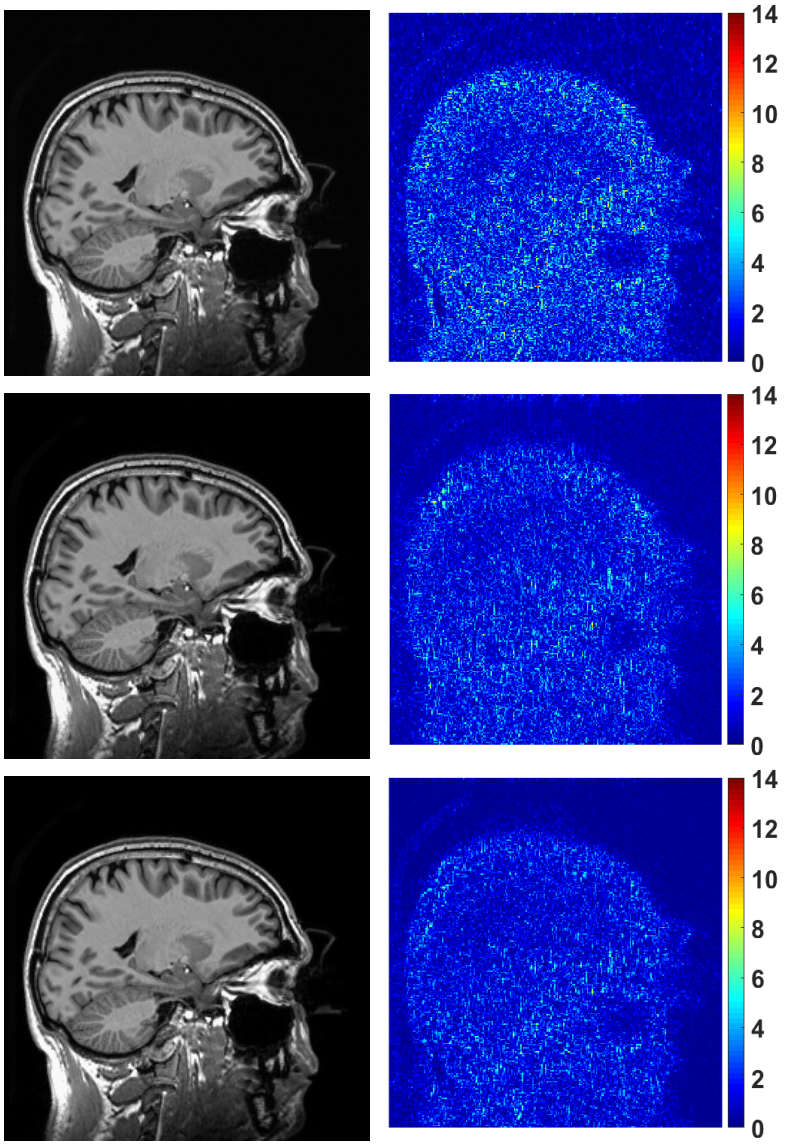



**Figure 4.10:** Reconstructed *sagittal1* image from 50% of random measurements. **First column**: WaTMRI (42.12 dB), LaSAL (44.27 dB) and LaSAL2 (45.37dB). **Second column**: Error in reconstruction for corresponding methods.

shows the results for random and radial sampling trajectories with the sampling rate of 30%. Three variants of pFISTA from [Liu 16] are tested, using contourlets, shift-invariant discrete wavelet (SIDWT) and patch-

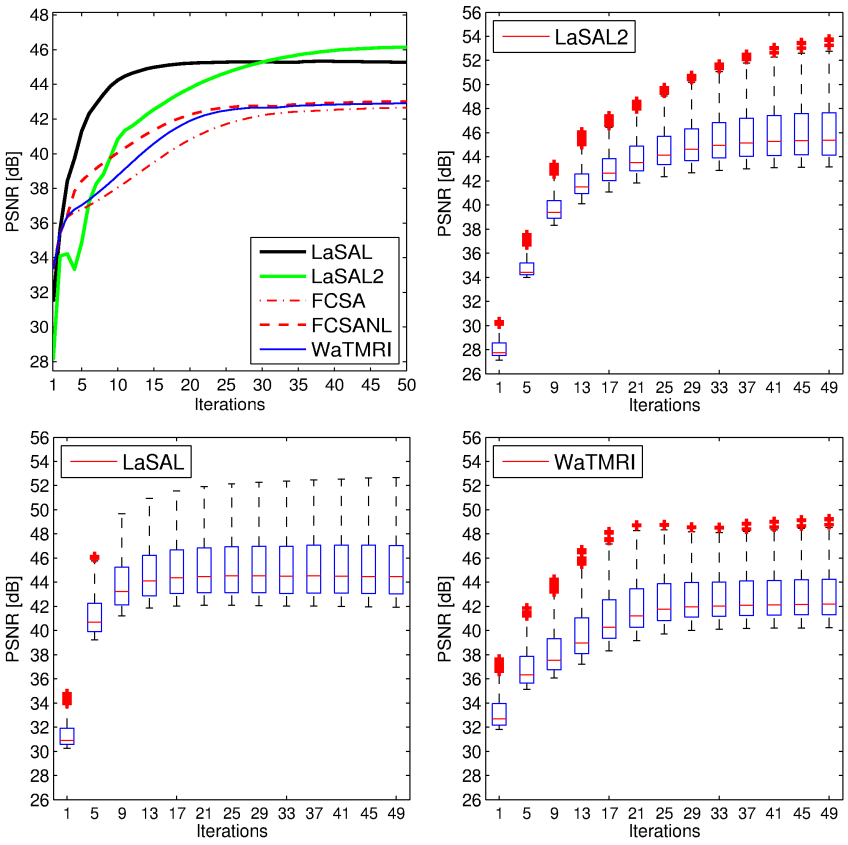



**Figure 4.11:** PSNR values obtained from 248 MRI *brain slices* from the first data set, with random sampling. Mean PSNR (**top left**) and the PSNR distribution for LaSAL2 (**top right**), LaSAL (**bottom left**) and WaTMRI (**bottom right**). The results are presented as a box plot: the edges of each box represent $25^{th}$ and $75^{th}$ percentile while the central mark (red line) in the box is median. The whiskers extend to the most extreme PSNR values which are not considered outliers while outliers are plotted separately with red crosses.

based directional wavelet (PBDW). We now used for LaSAL and LaSAL2 fewer shearlet bands (8, 4 and 2 per scale), resulting in comparable or smaller processing times with pFISTA, PANO, and DLMRI. For the radial trajectory, the best performing variant of pFISTA gives a similar (slightly better) results than LaSAL, but LaSAL2 clearly yields the higher PSNR value. With random sampling, both LaSAL and LaSAL2 outperform clearly all the variants of pFISTA, and LaSAL2 is again the best performing method. Moreover, in all the cases LaSAL reached the

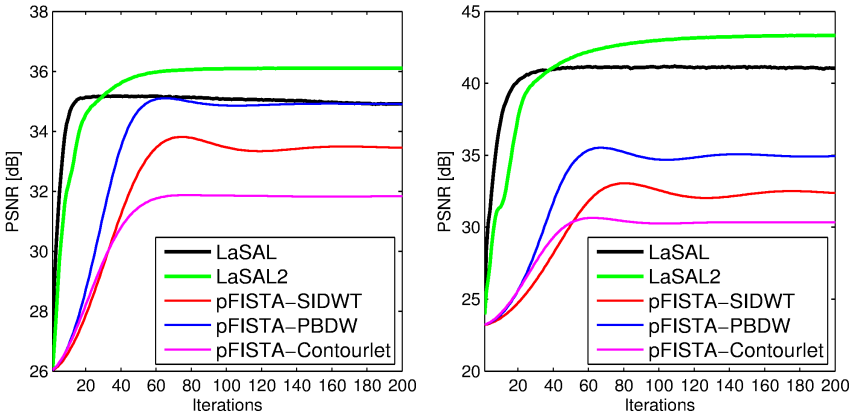



**Figure 4.12:** Comparison with pFISTA [Liu 16] on the test image *axial1*. **Left**: radial sampling (maximum PSNR values reached in LaSAL – 27s; LaSAL2 – 51 s; pFISTA-PBDW – 47 s;). **Right**: random sampling (maximum PSNR values reached in LaSAL – 39 s; LaSAL2 – 78 s; FISTA-PBDW – 58 s).

maximum PSNR faster than pFISTA (see the caption of Fig. 5.3).

**Table 4.1:** Comparison with PANO [Qu 14] and DLMRI [Ravishankar 11]

| ***axial2***, random lines 40% | | | ***axial3***, radial 14% | | |
|---|---|---|---|---|---|
| Method | PSNR [dB] | Time [s] | Method | PSNR [dB] | Time [s] |
| LaSAL | **43** | 111 | LaSAL | 36.6 | 507 |
| LaSAL2 | **45** | 105 | LaSAL2 | **39.4** | **231** |
| PANO | 41.3 | **74** | DLMRI | 37.5 | 763 |

For comparisons with PANO [Qu 14] and DLMRI from [Ravishankar 11], we used the images from the corresponding papers (*axial2* and *axial3* from Fig. 5.3, resp.), the sampling trajectories that were used in the corresponding papers as indicated in Table. 4.1, and the original publicly available codes. The resulting PSNR and processing times are listed in Table. 4.1. LaSAL2 outperforms both PANO and DLMRI method on their respective test data. LaSAL2 was somewhat slower than PANO and significantly (more than three times) faster than DLMRI. We also compared our approach to [Qu 10] and [Pejoski 15] on the test data from the original papers. Compared to the reported results in [Qu 10], LaSAL yields an improvement of nearly 2 dB, taking approximately the

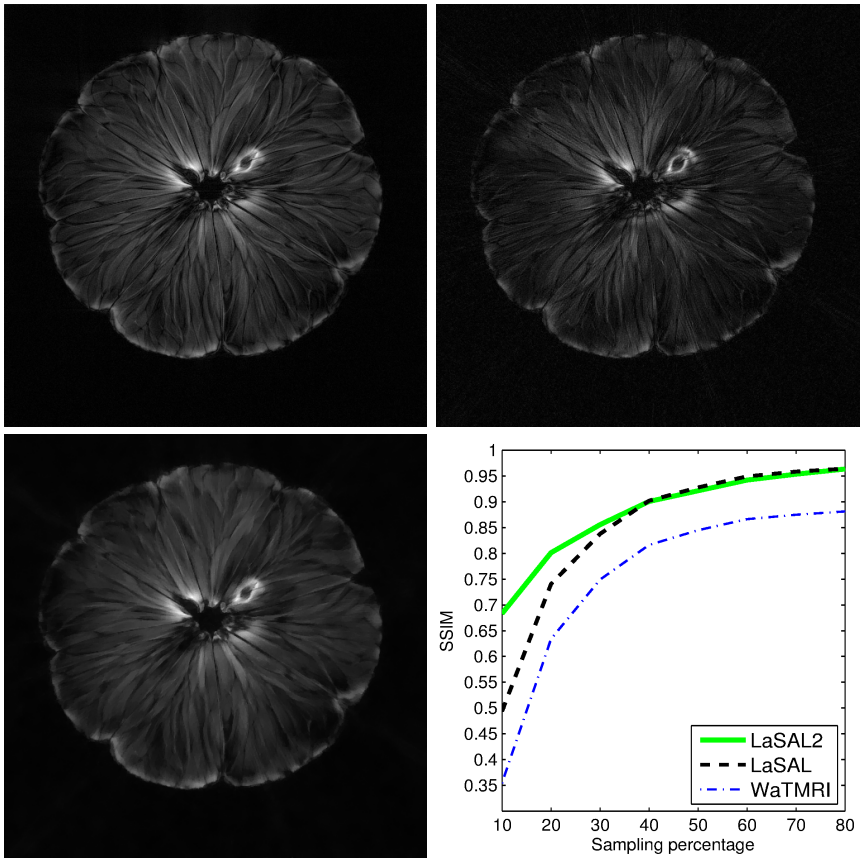



**Figure 4.13:** Reconstructions of the radially sampled *pomelo*. **Top left**: reconstructed from all available data with the conjugate gradient algorithm (reference image). **Top right**: WaTMRI reconstruction from 20% samples, SSIM = 0.65. **Bottom left**: LaSAL2 reconstruction from the same 20% samples, SSIM = 0.80. **Bottom right**: SSIM values for different sampling rates.

same time and LaSAL2 an improvement of more than 3 dB. The source code of [Pejoski 15] was unavailable, but compared to the reported results from this work, LaSAL yields similar or slightly better signal to noise ratio and LaSAL2 yields an improvement of 2.5 dB.

### 4.5.4 Experiments on radially sampled data

Here we perform experiments on a data set acquired with radial sampling in the $k$-space — an MRI scan of a *pomelo*, obtained from the BioImaging Lab at the University of Antwerp. The data consist of 1608 radial

lines, each with 1024 samples which are not necessary on the Cartesian grid. We form under-sampled versions by leaving out some of the radial lines. In particular, we aim to implement undersampling based on the golden ratio profile spacing [Winkelmann 07], which guarantees a nearly uniform coverage of the space for an arbitrary number of the remaining radial lines. Starting from an arbitrarily selected radial line, each next line is chosen by skipping an azimuthal gap of 111.246°. In practice, we cannot always achieve this gap precisely (since we have a finite, although a large, number of lines to start with). Therefore we choose the nearest available radial line relative to the position obtained after moving. Since we deal here with non-uniformly sampled $\not{k}$-space data, we need to employ the non-uniform FFT procedures [Fessler 03], which are commonly used in MRI reconstruction and readily available. In the reconstruction, we include weights on non-uniform measurements based on an area of Voronoi cells around each sample point. In [Rasche 99] is reported that using Voronoi weights as a measure of the local sampling density is very reliable. The three reference methods (WaTMRI, FCSA, and FCSANL) give similar results on *pomelo* image, so we choose for comparison WaTMRI. Fig. 4.13 shows visual comparison and SSIM values for LaSAL2 and WaTMRI. At sampling rates up to 30%, LaSAL2 reaches the highest SSIM, while for higher sampling rates it yields the same SSIM scores as LaSAL. For all sampling rates, both LaSAL and LaSAL2 outperform WaTMRI. At the highest sampling rates, WaTMRI's poor performance (SSIM is less than 0.9) is due to the employed regularization in WaTMRI which is related to penalization of the wavelet quadtree groups among the image coefficients. In this way, all coefficients inside the quadtree group are equally penalized which might lead to the loss in signal structure.

### 4.5.5 Convergence

The optimization problem that our method solves is non-convex. For a similar non-convex problem with MRF regularization, the authors in [Cevher 10b] argued that a local optimum can be efficiently obtained by applying alternating minimization. The same argument holds for our method. Although we cannot provide a theoretical proof of convergence, we provide solid empirical proof of convergence through simulation with different images and different trajectories. The experiments were conducted on various images using radial, spiral and random trajectories.

Fig. 4.14 shows the results for two different images and different

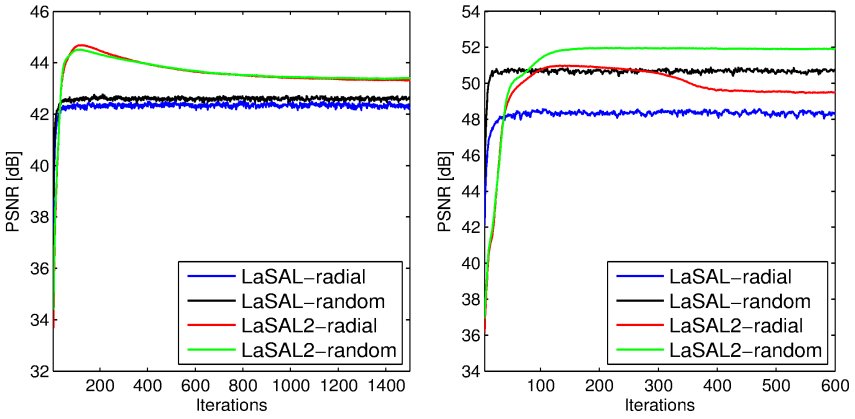



**Figure 4.14:** Experimental evaluation of the stability of the proposed methods on different test images. **Left**: *sagittal1*, 50% sampling. **Right**: *axial2*, 48% sampling.

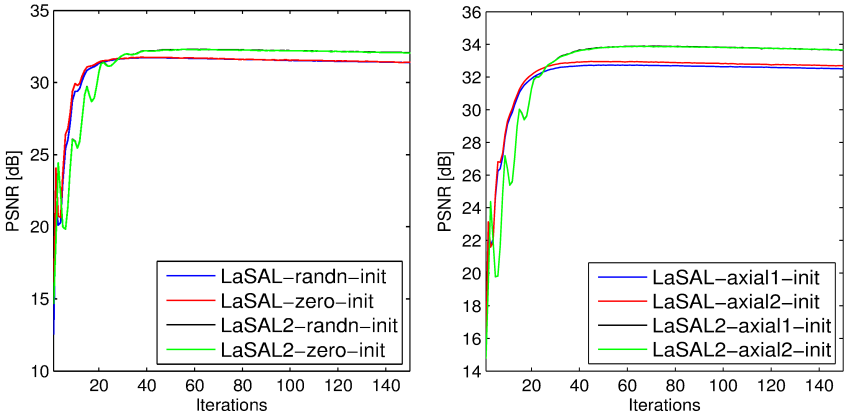



**Figure 4.15:** Influence of the initialization on the reconstruction performance illustrated on reconstructions of *saggital1* from 20% of measurements. **Left**: initializations with zero-image and random noise; random trajectory. **Right**: initializations with different MRI images (*axial1* and *axial2*); radial trajectory.

sampling trajectories. It can be observed that both LaSAL and LaSAL2 reach stable PSNR for all trajectories. In the case of LaSAL, only negligible oscillations persist around the converged value, while in the case of LaSAL2 no oscillations are observed. Changing the parameters of the MRF model can result in a higher maximum PSNR, at the cost of a less stable convergence.

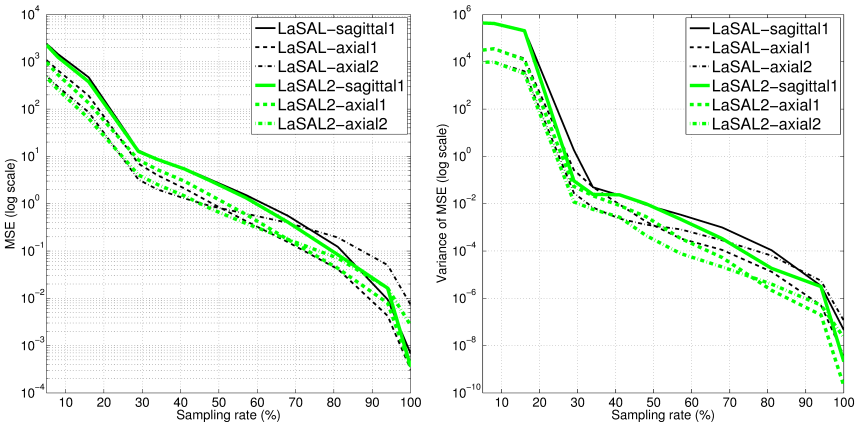



**Figure 4.16:** Experimental evaluation of the finite sample convergence of the proposed methods. MSE (left) and its variance (right) in reconstruction trials with three images at different sampling rates with random trajectories. For each sampling rate 50 experiments were conducted with different realizations of random trajectories and averaged MSE and its variance are plotted.

We also investigated the effect of initializing the reconstruction differently: with a zero image, with a white Gaussian noise image (zero mean, standard deviation 50) and with another MRI image as it is illustrated in Fig. 4.15. In the case of random noise initialization, we run 10 experiments and averaged results. The evolution of PSNR per iteration, after some initial iterations, practically does not depend on the initialization. We obtain similar results when initializing the reconstruction with an MRI image that is different from the one being reconstructed (see the diagram on the right of Fig. 4.15). In all our experiments, LaSAL and LaSAL2 reached their stable PSNR values that did not depend on the initial image.

Next, we analyze consistency of the proposed estimators. Estimation (reconstruction) of the original image from undersampled measurements is statistically consistent if the probability of reconstructing the true image converges to 1 as the number of measurements tends to infinity:

$$\lim_{n\to\infty} Pr(|T(\mathbf{y}_n) - \mathbf{x}| < \epsilon) = 1 \tag{4.35}$$

where $T$ denotes estimator, $n$ the number of samples in the measurements vector $\mathbf{y}$ and $\mathbf{x}$ the ground truth. Our proposed estimators LaSAL and LaSAL2 alternate between two minimization problems in an iterative procedure for image reconstruction. The first problem is the minimiza-

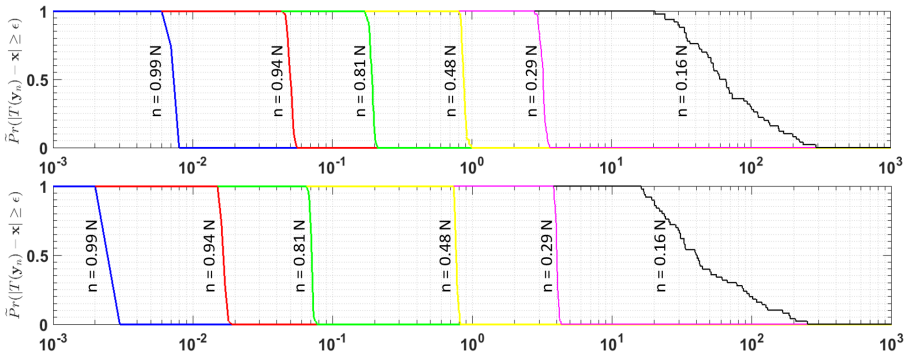



**Figure 4.17:** Empirically estimated probability $\widetilde{Pr}(|T(\mathbf{y}_n) - \mathbf{x}| \geq \epsilon)$ for various values of $\epsilon$ and different number of measurements $n = SP \times N$, where $SP$ denotes the sampling rate and $N$ is the size of the ideal image $\mathbf{x}$. Top: reconstructions with LaSAL (on *sagittal1*) Bottom: the corresponding results for LaSAL2.

tion of an energy function composed of a data fitting term and a prior energy term, expressed as the energy of an Ising MRF model. This minimization results in estimated support of the signal in a transformation domain. A detailed analysis of Gibbs-Markov random field models including statistical consistency of minimum contrast estimators employing these models is provided in [Gaetan 10]. The second problem is an objective function minimization that estimates the signal, constrained to the particular domain (signal space) imposed by the previously estimated signal support. Since this particular objective function is convex, its consistency is trivially proven.

LaSAL and LaSAL2 procedures alternate between the two estimators, inferring jointly the signal and its support in a transform domain. Proving the consistency of each estimator separately does not lead directly to the consistency proof of the joint estimator; the conditions are studied in literature [Niesen 09], but such a rigorous analysis exceeds the scope of this work. We provide a finite sample convergence analysis of the joint estimator using an experimental setup similar to the one in [Loh 17]. This experiment evaluates statistically image reconstruction quality as a function of an increasing number of measurements. At each of the considered sampling rates, we perform reconstructions over 50 realizations of randomly generated acquisition trajectories (where each reconstruction contains over 65000 pixels for a 256×256 image) and we record the averaged mean squared error (MSE) and its variance over all the realizations. The evaluation of MSE in this setting is commonly

motivated in the statistical literature by Chebychev's inequality, from which it follows that:

$$Pr(|T(\mathbf{y}_n) - \mathbf{x}| \geq \epsilon) \leq \frac{E((T(\mathbf{y}_n) - \mathbf{x})^2)}{\epsilon^2} \tag{4.36}$$

Fig. 4.16 shows the results obtained for three different input images. Relatively high values of variance at small sampling rates (less than 30%) can be attributed to the fact that random trajectories may miss (almost completely) or not the lowest frequency components which are essential for the quality of reconstruction. We conclude that as the number of measurements increases, the MSE and its variance decrease and tend to zero, as expected.

It is interesting to examine also the empirical estimates of the probabilities $Pr(|T(\mathbf{y}_n) - \mathbf{x}| \geq \epsilon)$, which can be obtained from the same experimental setup. For each $n$, we find an empirical estimate $\widetilde{Pr}(|T(\mathbf{y}_n) - \mathbf{x}| \geq \epsilon)$ as the fraction of the total number of experiments for which the absolute error of the reconstruction was exceeding $\epsilon$. We illustrate these empirical probabilities for one of the test images in Fig. 4.17. The diagrams show that at very high sampling rates the empirical probability $\widetilde{Pr}(|T(\mathbf{y}_n) - \mathbf{x}| \geq \epsilon)$ indeed tends to zero for $\epsilon > 0.008$ (LaSAL) or $\epsilon > 0.003$ (LaSAL2) on grayscale images in the range [0,255]. These results indicate that the proposed algorithm reliably converges to solutions that lie within a standard deviation that can be ignored safely in any practical application.

## 4.6 FCLaTV

Obtained results in the reconstruction of MR images using LaSAL and LaSAL2 methods in the previous section showed substantial improvement in image quality compared to other methods. The image structure (edges and texture) is very well preserved and the overall contrast in the image is not reduced. This was mostly accomplished using a hard thresholding based rule on image coefficients which is derived using the adopted MRF signal prior.

Since MR images are not completely sparse (they can be approximated as sparse), the signal energy is not only concentrated in the sparse coefficients with the highest magnitude. The hard-thresholding rule keeps coefficients unchanged on those positions that are indicated by the estimated signal support. The locations of important image structures are revealed by estimated support not only based on a high coefficient's magnitude on those locations. Therefore, among the unchanged

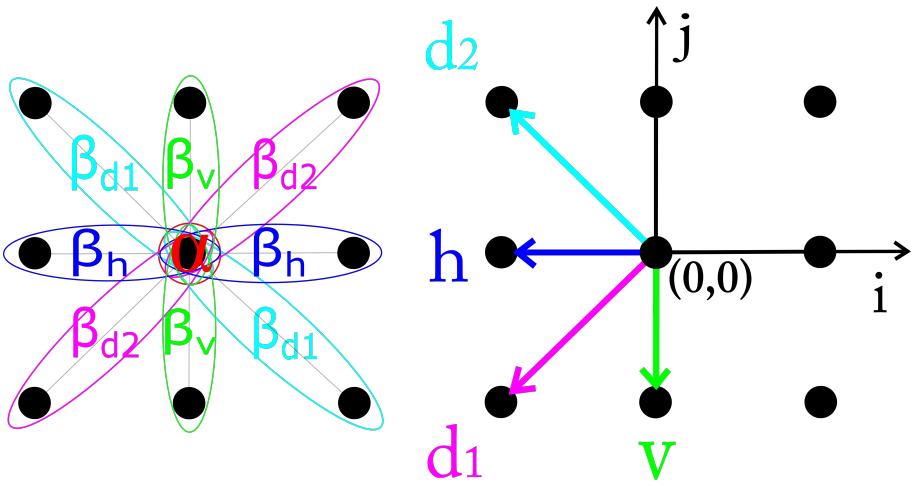



**Figure 4.18: Left:** Single and pair-site cliques for the second-order neighborhood with parameters. **Right:** Directions used in calculation of $R_t$.

image coefficients there can be ones with low magnitude where the signal strength is disturbed more by the present noise compared to ones that have high magnitude. This leads to the conclusion that an additional or modification of the existing rule for coefficients denoising should be proposed that would take into account the magnitude of the coefficients along with their support.

In this subsection we propose the optimization based method which uses soft-thresholding regularization rule derived taking into account the estimated signal support. We employ a MRF model for signal support with pair-site cliques which is upgrade of the model utilized in our previous work [Pižurica 11,Panić 16a,Panić 17a]. A different interaction coefficients for cliques of different orientations is allowed, leading to anisotropy property of the MRF model, which better characterized the actual subband statistics. The energy function of the adopted anisotropic MRF model is

$$H(\mathbf{s}) = \sum_i \alpha s_i + \sum_{\langle i,j\rangle \in \mathcal{C}} \beta_o s_i s_j \tag{4.37}$$

where $\alpha$ expresses the a priori preference for labels of one type over the other and $\beta_o$ the interaction strength for cliques with orientation $o \in \{h, v, d1, d2\}$ shown in Fig. 4.18. The estimation of the MRF parameters $(\beta_o, \alpha)$ is treated in the following subsection 4.6.1.

We recall the problem formulation in (4.33) with two regularization functions (one with MRF prior and second based on TV norm)

from subsection 4.4, here slightly different formulated

$$\min_{\mathbf{x}\in\mathbb{C}^N} f(\mathbf{x}) + \phi_{\text{MRF}}(\mathbf{P}\mathbf{x}) + \|\mathbf{x}\|_{\text{TV}} \tag{4.38}$$

where $f(\mathbf{x})$ corresponds to data fidelity term which instead of indicator function is expressed through $\ell_2$ norm loss like in (3.24). The first regularization function is given in the form of optimization problem

$$\phi_{\text{MRF}}(\mathbf{P}\mathbf{x}) = \max_{\mathbf{s}} p_{\boldsymbol{\theta}|\mathbf{S}}(\boldsymbol{\theta} \mid \mathbf{s}) P_{\mathbf{S}}(\mathbf{s}) = \min_{\mathbf{s}} -\big(\log p_{\boldsymbol{\theta}|\mathbf{S}}(\boldsymbol{\theta}|\mathbf{s}) + \log P_{\mathbf{S}}(\mathbf{s})\big) \tag{4.39}$$

where $P_{\mathbf{S}}(\mathbf{s})$ is defined as MRF anisotropic Ising-model with energy function given in (4.37) while $p_{\boldsymbol{\theta}|\mathbf{S}}(\boldsymbol{\theta}|\mathbf{s})$ has the same form as in previous proposed methods (see subsection 4.2.4).

To solve the problem (4.38) we adopt the idea of fast composite splitting from [Huang 11a] and extend this algorithm to deal with the MRF-based regularization. The key difference is that our model involves $\phi_{\text{MRF}}$ instead of $\ell_1$-norm in [Huang 11a]. Hence, we have to solve the proximal map $\hat{\boldsymbol{\theta}} = \boldsymbol{\Psi}_{\phi_{\text{MRF}}}(\mathbf{P}\mathbf{x}_g; \mu)$:

$$\underset{\boldsymbol{\theta}}{\text{argmin}} \left( \phi_{\text{MRF}}(\boldsymbol{\theta}) + \frac{1}{2\mu} \|\boldsymbol{\theta} - \mathbf{P}\mathbf{x}_g\|_2^2 \right). \tag{4.40}$$

Note that, in notation $\boldsymbol{\Psi}_{\phi_{\text{MRF}}}(\mathbf{P}\mathbf{x}_g; \mu)$, $\phi_{\text{MRF}}$ is the function for which the proximal map is calculated, and $\boldsymbol{\theta}_g = \mathbf{P}\mathbf{x}_g$ is the argument at which it is evaluated. Since the evaluation of this proximal map is very hard, we adopt a suboptimal, block-coordinate approach explained in the following.

The key novel ingredient of the proposed algorithm is a computationally efficient approximation of the proximal map in (4.40) for a fixed $\mathbf{x}_g$. From (4.39), it is clear that the minimization in (4.40) needs to be carried out jointly w.r.t. $\boldsymbol{\theta}$ and $\mathbf{s}$ in order to evaluate (4.40) exactly. We adopt a suboptimal yet computationally efficient approach where we first minimize the second term in (4.39) w.r.t. $\mathbf{s} \in \{-1, 1\}^D$ for a fixed $\boldsymbol{\theta}$; then we fix the obtained $\hat{\mathbf{s}}$ and minimize the objective in (4.40) w.r.t. $\boldsymbol{\theta}$. For the former minimization step, we adopt the same notation as in previous proposed methods $\hat{\mathbf{s}} = \mathit{MAP\text{-}support}\{\boldsymbol{\theta}\}$. The step is done via the Metropolis sampler procedure using a "warm-start" initial $\mathbf{s}$ as in [Panić 16a, Panić 17a]. The latter minimization (w.r.t. $\boldsymbol{\theta}$ for a fixed $\hat{\mathbf{s}}$), as shown here, can be done in closed-form and leads to a novel soft-thresholding operation. This is accomplished by solving the following

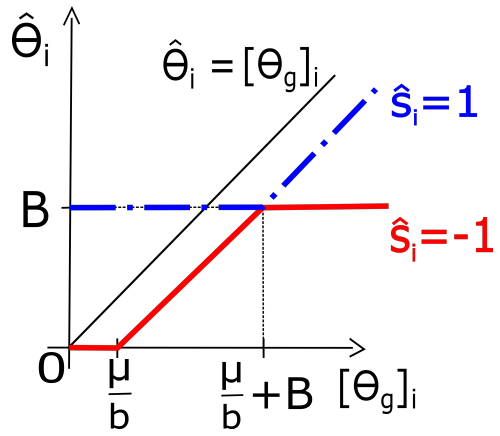



**Figure 4.19:** Soft-thresholding rules for image coefficients $[\theta_g]_i$ based on the estimated $\hat{s}_i = \{1, -1\}$.

minimization problem

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \underbrace{\sum_i -\log p(\theta_i|\hat{s}_i) + \frac{1}{2\mu}(\theta_i - [\theta_g]_i)^2}_{\phi_{\hat{\mathbf{s}}}(\boldsymbol{\theta})}. \tag{4.41}$$

Equation (4.41) is derived from (4.40) using a simple algebraic manipulation and omitting the terms that do not depend on $\boldsymbol{\theta}$. With the analytic form of $p_{\Theta_i|S_i}(\theta_i|s_i)$ for $\hat{s}_i = \{-1, 1\}$ (see subsection 4.2.4), the closed form solution for each single entry of $\hat{\mathbf{s}}$ (it turns out that the solution decouples entry-wise) is derived in (4.42) (index $i$ in the equation is omitted for notation simplicity). The solution (4.42) is illustrated in Fig. 4.19 on range $[0, \infty)$ due to its (odd) symmetry about the origin.

$$\begin{aligned}
\hat{\theta}_{\hat{s}=1} &= \begin{cases} B \cdot \operatorname{sgn}(\theta_g), & |\theta_g| \leq \frac{\mu}{b} + B \\ \theta_g - \frac{\mu}{b} \cdot \operatorname{sgn}(\theta_g), & |\theta_g| > \frac{\mu}{b} + B \end{cases} \\
\hat{\theta}_{\hat{s}=-1} &= \begin{cases} 0, & |\theta_g| \leq \frac{\mu}{b} \\ \theta_g - \frac{\mu}{b} \cdot \operatorname{sgn}(\theta_g), & \frac{\mu}{b} < |\theta_g| \leq \frac{\mu}{b} + B \\ B \cdot \operatorname{sgn}(\theta_g), & |\theta_g| > \frac{\mu}{b} + B \end{cases}
\end{aligned} \tag{4.42}$$

The described block-coordinate cyclic procedure should in principle proceed in several iterative rounds. As we demonstrate here numerically, it is sufficient to perform a single cycle. Once the approximation of

**Algorithm 13** FCLaTV

**Input:** $k = 1, \mu, \tau_1, \tau_2, t^{\{1\}} = 1, \mathbf{y}, \mathbf{r}^{\{1\}} = \mathbf{x}^{\{0\}}$

1: **repeat**
2: $\quad \mathbf{x}_g = \mathbf{r}^{\{k\}} - \mu \mathbf{A}^H(\mathbf{A}\mathbf{r}^{\{k\}} - \mathbf{y})$
3: $\quad [\phi_{\hat{\mathbf{s}}}, \hat{\mathbf{s}}] = \mathit{MAP\text{-}support}\{\mathbf{P}\mathbf{x}_g\}$
4: $\quad \mathbf{x}_1 = \mathbf{P}^H(\mathbf{prox}_{\mu}(\tau_1 \phi_{\hat{\mathbf{s}}})(\mathbf{P}\mathbf{x}_g))$
5: $\quad \mathbf{x}_2 = \mathbf{G}^H(\mathbf{prox}_{\mu}(\tau_2 \psi_{\mathrm{TV}})(\mathbf{G}\mathbf{x}_g))$
6: $\quad \mathbf{x}^{\{k\}} = (\mathbf{x}_1 + \mathbf{x}_2)/2$
7: $\quad t^{\{k+1\}} = \left(1 + \sqrt{1 + 4(t^{\{k\}})^2}\right)/2$
8: $\quad \mathbf{r}^{\{k+1\}} = \mathbf{x}^{\{k\}} + \frac{t^{\{k\}}-1}{t^{\{k+1\}}}(\mathbf{x}^{\{k\}} - \mathbf{x}^{\{k-1\}})$
9: $\quad k = k + 1$
10: **until** some stopping criterion is satisfied
11: **return** $\mathbf{x} = \mathbf{x}^{\{k\}}$

the proximal map in (4.40) is ready, we incorporate it in a fast-splitting framework akin to this in [Huang 11a]. That is, we first carry out a gradient-like step on $\mathbf{x}$ w.r.t. data fidelity term. Then, we perform in parallel the proximal maps that correspond to the two regularization functions in (4.38). The two outputs of the proximal mappings are then simply averaged and fed into a Nesterov-acceleration-like step. More iterations of calculating the proximal maps and their average can be made to a refined approximation of the proximal map in (4.40), which it turns out that doesn't bring much improvement in reconstruction in our methods as well as in the FCSA method from [Huang 11b].

The motivation for the parallel-proximal-then-average approach is computational efficiency, as carrying out a joint proximal map (even approximate) w.r.t. $\phi_{\mathrm{MRF}}$ and $\psi_{\mathrm{TV}}$ would be very hard. The motivation for the Nesterov-acceleration step is to further speed-up the method which is reported for methods in [Huang 11b, Huang 11a]. The overall method is presented in Algorithm 13. Note that algorithm steps 3-4 are the single block-coordinate cycle to approximate the operation in (4.39). After the estimation of signal support $\hat{\mathbf{s}}$ in step 3, the $\phi_{\hat{\mathbf{s}}}$ term in (4.41) is calculated which further results in two soft-thresholding rules derived in (4.42). The parameters $\tau_1, \tau_2, \mu$ for simplicity are all set to 1, although these are not the optimal values unless it is other stated.

We refer to the proposed method as FCLaTV (Fast Composite Lattice and TV regularization). A version without acceleration CLaTV can also be used, and it is obtained after omitting steps 7 and 8 from FCLaTV and replacing $\mathbf{r}^{\{k\}}$ in step 1 with $\mathbf{x}^{\{k\}}$. Extensive numerical

studies demonstrate that Algorithm 13 always converges. The actual rigorous convergence analysis is left for future work.

### 4.6.1 Parameter estimation for the anisotropic MRF prior

Here we propose a data-driven approach for specifying the parameters $\{\alpha, \beta_h, \beta_v, \beta_{d1}, \beta_{d2}\}$ of the MRF model. The core idea is to relate the parameters of the prior $P(\mathbf{s})$ for the support of $\boldsymbol{\theta}$ to some measurable characteristics of the observed $\boldsymbol{\theta}$. The representation coefficients $\boldsymbol{\theta}$ are re-estimated in each iteration of Algorithm 13 and so are the MRF parameters.

The four interaction coefficients $\{\beta_h, \beta_v, \beta_{d1}, \beta_{d2}\}$ represent the clustering strength of the subband coefficient labels in the corresponding four directions. We reason that the interaction strength in a particular direction should be proportional to the correlation strength of the corresponding representation coefficients $\boldsymbol{\theta}$ in that direction. In order to reduce the effect of noise we squared the coefficients and account only on those that were marked as significant by the initial $\mathbf{s}$ in the given iteration. In what follows, $\boldsymbol{\theta}^{\mathrm{S}}$ denotes a subband in which all the coefficients that were not selected as significant by $\mathbf{s}$ are set to zero ($\theta_i^{\mathrm{S}} = \theta_i$ if $s_i = 1$ and $\theta_i^{\mathrm{S}} = 0$ otherwise). To express two-dimensional (2D) correlation, we need to revert to 2D spatial indices $(k, l)$. Let $r_{(i,j)}$ denote the correlation coefficient for squared $\boldsymbol{\theta}^{\mathrm{S}}$ corresponding to the spatial shift $(i, j)$:

$$r_{(i,j)} = \sum_k \sum_l \left(\theta^{\mathrm{S}}(k,l)\theta^{\mathrm{S}}(k+i, l+j)\right)^2 \tag{4.43}$$

We take four correlation coefficients corresponding to the smallest possible spatial shifts in the corresponding directions, indicated by arrows in Fig. 4.18. These are: $r_{(-1,-1)}$, $r_{(-1,0)}$, $r_{(-1,1)}$ and $r_{(0,-1)}$. Note that by symmetry $r_{(-1,-1)} = r_{(1,1)}$, $r_{(-1,0)} = r_{(1,0)}$, $r_{(-1,1)} = r_{(1,-1)}$ and $r_{(0,-1)} = r_{(0,1)}$. The mapping $t = 3i + j + 4$ translates these index pairs into a single index $t \in \{0, 1, 2, 3\}$. With $R_t = r_{(i,j)}$ we have that $R_0 = r_{(-1,-1)}$ is the correlation coefficient in the 45° ($d_1$-direction), $R_1 = r_{(-1,0)}$ the horizontal, $R_2 = r_{(-1,1)}$ the 135° ($d_2$-direction) and $R_3 = r_{(0,-1)}$ in the vertical direction. Now we can specify the four inter-

action coefficients as normalized correlation coefficients of $\boldsymbol{\theta}^{\mathrm{S}}$:

$$\begin{aligned} \beta_{d1} &= \frac{R_0}{\|\mathbf{R}\|_2}, \quad \beta_h = \frac{R_1}{\|\mathbf{R}\|_2} \\ \beta_{d2} &= \frac{R_2}{\|\mathbf{R}\|_2}, \quad \beta_v = \frac{R_3}{\|\mathbf{R}\|_2} \end{aligned} \tag{4.44}$$

The normalization by $\ell_2$-norm here is optional (as only the relative values of the MRF parameters with respect to each other actually matter) but we find it convenient in practice to have these parameters in the range [-1,1] as it is now guaranteed. We also tested $\ell_1$ and $\ell_\infty$ for the purpose of this normalization, but $\ell_2$ led to best performances in our experiments.

It still remains to specify the parameter $\alpha$, which represents *a priori* preference for one type of labels ($-1$ or $+1$) over the other. With $\alpha = 0$ both labels are a priori equally likely and as $\alpha$ increases in magnitude the more preference goes to one of these labels. For $\alpha > 0$, the labels $-1$ will be favoured, which means that significant coefficients (labelled by $+1$) will be sparse.We specify $\alpha$ as the mean energy of the coefficients in $\boldsymbol{\theta}^{\mathrm{S}}$ relative to the energy of the largest coefficient in that subband:

$$\alpha = \frac{1}{N} \frac{\|\boldsymbol{\theta}^{\mathrm{S}}\|_2^2}{\left(\|\boldsymbol{\theta}^{\mathrm{S}}\|_{\max}\right)^2} \tag{4.45}$$

Note that we have omitted the iteration indices for compactness. In fact, we have sequences $\mathbf{s}^{(k)}$, $\boldsymbol{\theta}^{\mathrm{S}(k)}$ and $\alpha^{(k)}$, $\beta_h^{(k)}$, $\beta_v^{(k)}$, $\beta_{d1}^{(k)}$, $\beta_{d2}^{(k)}$ that get improved through iterations $k$.

## 4.7 Experiments and Discussion

In the experimental evaluation, we use the same data set acquired on a Cartesian grid, explained in subsection 4.5, which is used in the testing of LaSAL and LaSAL2 methods. First we tested the proposed methods for reconstruction of *sagittal1* slice, shown in Fig. **??**, for different sampling rates (SR) from radially undersampled measurements. Then we evaluate reconstructions of all 248 slices from this dataset for the particular SR of 48%. We conducted experiments with the non-Cartesian sampled data from the pomelo dataset, which is explained in subsection 4.5, with the same experimental setup used for testing LaSAL and LaSAL2. A complex image reconstruction (its real and imaginary part) is also considered with the *axial1* slice from [Liu 16] (its magnitude is presented in Fig. 5.3). Measurements in this experiment are obtained

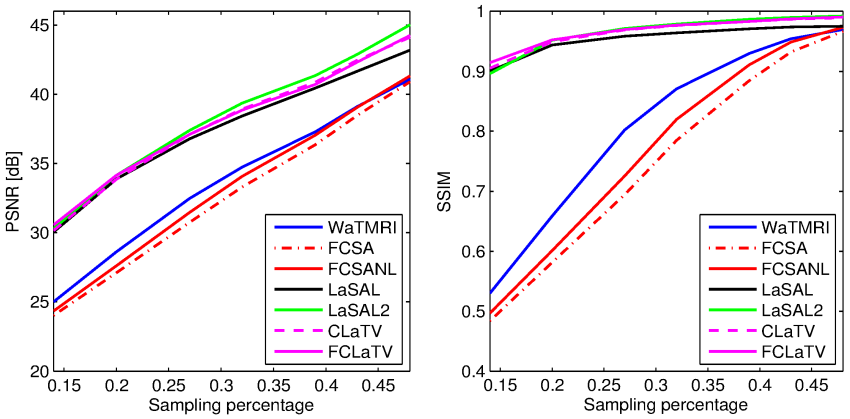



**Figure 4.20:** PSNR and SSIM for the reconstructions of the test image *sagittal1* at different sampling rates.

using radial and random sampling trajectories. In all our experiments we used the non-decimated wavelet transform with 3 scales. For comparison, we report the results of LaSAL and LaSAL2 from [Panić 17a], FCSA [Huang 11b], FCSANL [Huang 12] and WaTMRI [Chen 14] with the original implementations. All these methods, except LaSAL, employ a compound regularization. In the following two subsections we divide experimental results according to the sampling grid used in measurement acquisition.

### 4.7.1 Data sets acquired on the Cartesian grid

We first tested our methods on data set which comprises T1 MRI images previously explained in subsection 4.5. Fig. 4.20 shows the Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM) for *sagittal1*, with radial trajectory and sampling rate (SR) ranging from 14% to 48%. The MRF-based methods LaSAL, LaSAL2, CLaTV, and FCLaTV achieve a consistent and significant improvement in PSNR (at some sampling rates more than 4 dB) compared to WaTMRI, FCSA, and FCSANL. The proposed methods CLaTV and FCLaTV outperform LaSAL and yield only slightly lower reconstruction PSNR and equally good SSIM as the best reference method LaSAL2. These results are achieved with an automatic estimation of MRF parameters and without tuning of regularization parameters (the $\mu$,$\tau_1$ and $\tau_2$ parameters are all set to 1).

We tested performances of the proposed methods through the

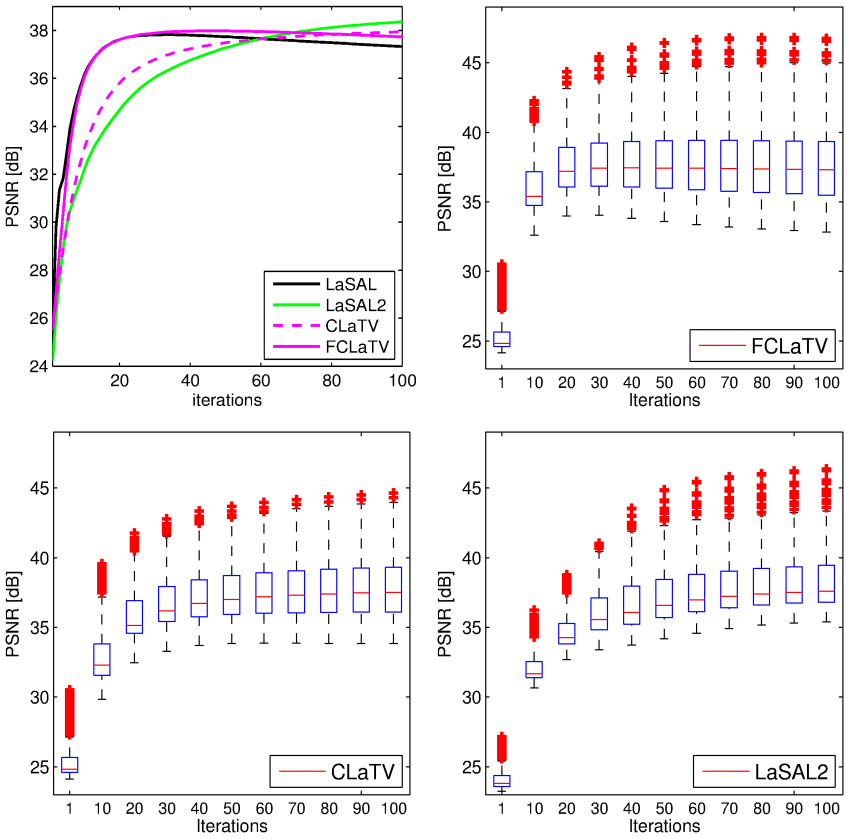



**Figure 4.21:** PSNR values obtained from 248 MRI *brain slices* from the first data set, with radial sampling (SR 25%). Mean PSNR (**Left**) and the PSNR distribution for FCLaTV (**right**). The result is presented as a box plot: the edges of each box represent $25^{th}$ and $75^{th}$ percentile while the central mark (red line) in the box is median. The whiskers extend to the most extreme PSNR values which are not considered outliers while outliers are plotted separately with red crosses.

reconstruction of all 248 MRI slices from the data set. Results are shown in Fig. 4.21 in terms of PSNR. We also provide distribution of PSNR values through iterations for CLaTV, FCLaTV and LaSAL and LaSAL2 methods. On average FCLaTV reaches the peak performance much before CLaTV and LaSAL2 and keeps it steady through iterations, unlike LaSAL. Although LaSAL2 achieved the highest PSNR on average (greater than CLaTV and FCLaTV in less than 1 dB) this is accomplished after 70 iterations which is more than 3 times later after

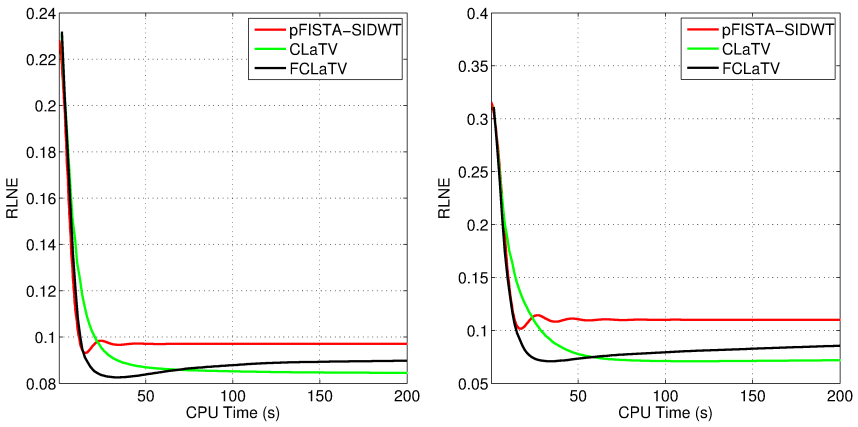



**Figure 4.22:** Reconstruction PSNR for the image *axial1*. **Left**: radial and **Right**: random trajectory with the same SR of 30%.

FCLaTV. CLaTV, FCLaTV, and LaSAL2 reached the same maximal median PSNR value of 37.5 dB through iterations. CLaTV and FCLaTV need fewer iterations than LaSAL2 to reach this value which is presented in Fig. 4.21 through the shift of PSNR distribution towards higher values in the first 20 iterations.

The proposed methods are tested for reconstruction of complex MRI *axial1* slice which real and imaginary part are shown in Fig. **??**. For the performance measure we use relative $\ell_2$ norm error (RLNE) which is defined as

$$e(\hat{\mathbf{x}}) = \frac{\|\hat{\mathbf{x}} - \mathbf{x}_r\|_2}{\|\mathbf{x}_r\|_2} \tag{4.46}$$

where $\mathbf{x}_r$ denotes reference image reconstructed from all measurements with inverse Fourier transform, while $\hat{\mathbf{x}}$ represents estimated image from the undersampled measurements. Steps 4 and 5 in both version of algorithm (with and without FISTA acceleration) which refer to proximal operators are simultaneously applied on real and imaginary part of temporary reconstructed image $\mathbf{x}_g$ obtained from step 2. We found that this adaptation of the algorithm achieves the best performances in case when reconstruction of complex images is considered.

We compared the proposed methods with the pFISTA method from [Liu 16] which uses the Shift Invariant Discrete Wavelet Transform (SIDWT) as a tight frame for sparse signal representation. The results are shown in Fig. 4.22 with respect to the amount of CPU time needed for the reconstruction. For both employed sampling trajectories,

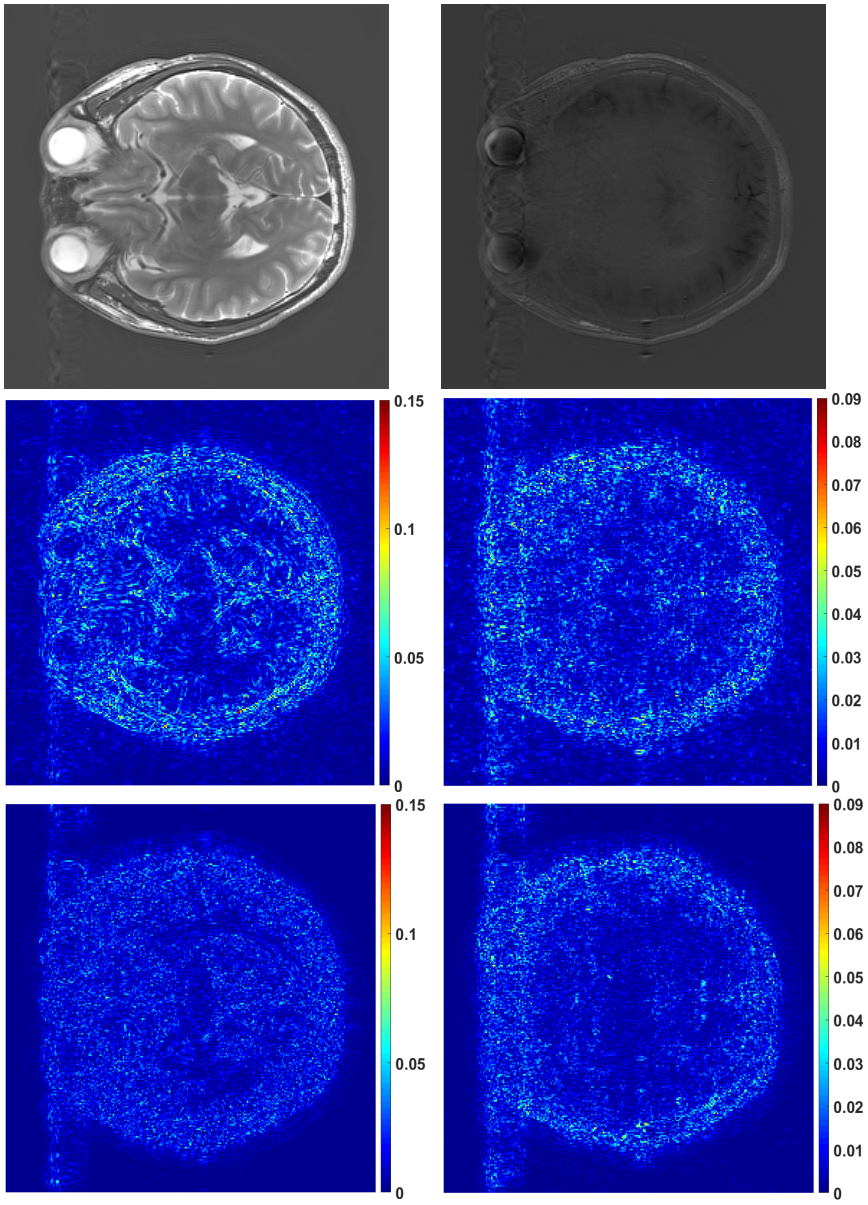



**Figure 4.23:** Reconstruction of the **Left column**: real and **Right column** imaginary part of the image *axial1* using random trajectory with the same SR of 30%. From top to bottom are given referent real and imaginary part of the image *axial1* and their corresponding reconstructions errors with pFISTA-SIDWT and FCSLa respectively.

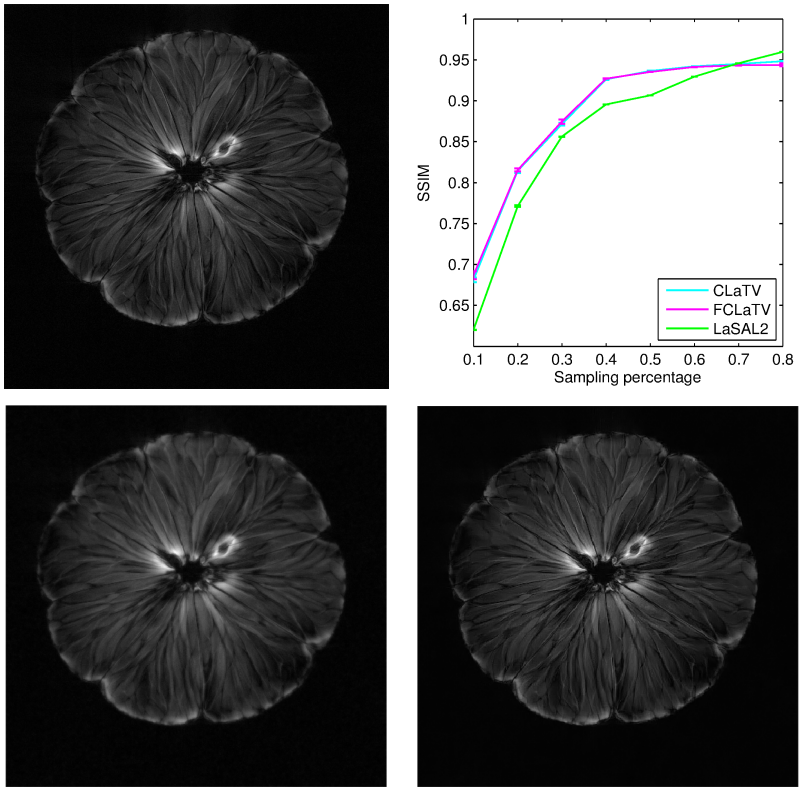



**Figure 4.24:** Reconstruction of pomelo fruit from radially sampled measurements. **The first row left to right**: reference image obtained by conjugate gradient method using all measurements, avarege SSIM for different sampling rate. **The second row left to right**: pomelo reconstructions using LaSAL2 and FCLaTV methods respectively from 40% of measurements.

CLaTV and FCLaTV outperform pFISTA-SIDWT in terms of RLNE measure. FCLaTV reaches the lowest value of RLNE in less than the 50s while CLaTV needs around 70s. The lowest RLNE values achieved with FCLaTV are 0.0826 and 0.0709 while for CLaTV are 0.0826 and 0.0709 for radial and random trajectory respectively. It is worth mentioning that usage of non-decimated wavelet transform significantly reduced computational cost per algorithm iteration due to fewer subbands are involved for support estimation compared to non-decimated shearlet transform. Visual comparison of reconstructed real and imaginary part of the *axial1* slice using FCLaTV and pFISTA-SIDWT methods is given in Fig. 4.23 with reconstruction errors. We can see that the image structure is better recovered with the FCLaTV method.

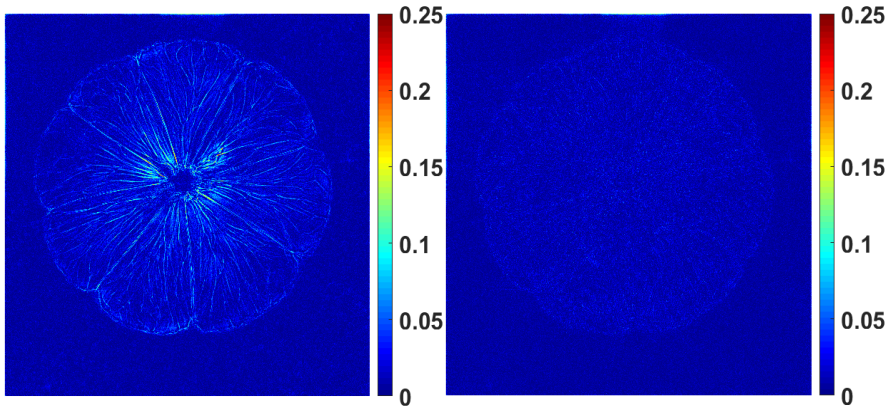



**Figure 4.25:** Coresponding reconstructions errors using LaSAL2 and FCLaTV methods respectively using 40% of measurements.

Obtained results of FCLaTV shown fast convergence (in less than 25 iterations) similar to the LaSAL method but without oscillatory behavior in later iterations which is the same property that LaSAL2 has shown. With this, we accomplish to have two desired method characteristics which are stability and fast convergence. This is accompanied by automatic parameter estimation which is very important for practical application and simplifies algorithm usage. The algorithm complexity is significantly reduced by using a non-decimated wavelet instead of shearlet transform. The obtained performances are in the range of the LaSAL2 method or slightly lower and still significantly higher than other state of the art methods for reconstruction.

### 4.7.2 Data sets acquired on non-Cartesian grid

We perform experiments on an MRI scan of the *pomelo*, the same dataset used in the testing of our previous proposed methods LaSAL and LaSAL2. The *pomelo* dataset and the same experimental setup that we used here were described in subsection 4.5.4. For the reference method we select LaSAL2 [Panić 17a]. During the reconstruction, we add on undersampled measurements a different amount of Gaussian noise with zero mean and the following standard deviations 0.01, 0.02, 0.03, 0.04 and 0.05. For each amount of added noise, we simulate different realizations of noise and perform 10 experiments. The overall performance is expressed in SSIM measure which is calculated first by averaging SSIM's across all 10 experiments for each particular noise level

and then by averaging obtained SSIM's values across all suggested noise levels. We present results in Fig. 4.24 with obtained mean SSIM value together with standard deviation bars to denote algorithm sensitivity to a different amount of noise. We empirically set $\tau_2$ to 140 and 130 for CLaTV and FCLaTV respectively in order to achieve the best algorithms performances. This is done due to the different dynamic range of [0,1] for the used reference image. Regularization parameters for the LaSAL2 method were the same as in subsection 4.5.4. As expected for all methods standard deviation of performances reduces with the increase of the sampling rate while the mean SSIM becomes greater. A relatively small sampling rate, up to 60% CLaTV and FCLaTV outperforms LaSAL2 which is also present in visual comparison of reconstructed images in the same Fig. 4.24. In general CLaTV and FCLaTV better recover structures in the image than LaSAL2 (i.e. the reconstructed image is less blurry) which is reflected in the reconstruction errors as well, presented in Fig. 4.25.

## 4.8 Conclusion

In this chapter, we proposed four optimization-based algorithms for MR image reconstruction which utilized MRF signal prior. For regularization steps in algorithms iterations, we derived two thresholding rules based on the estimated signal support in the signal transformation domain. Two versions of the Ising model, isotropic and anisotropic are considered for the model of the signal support, the later with the automatic procedure for MRF parameter estimation. The MRF based regularizations are further improved by involving the TV norm regularization in algorithm development. We considered two iterative frameworks CSALSA and FCSA in which we incorporated the proposed regularizations based on the MRF model alone or together with TV norm. Both iterative frameworks analyze the possibility of including more than one regularization term in the objective function and suit well for the problem considered here in terms of complexity and parameters involved. Using the CSALSA framework we derived two methods LaSAL and LaSAL2, while with the FCSA framework we proposed CLaTV and FCLaTV. Among developed methods, LaSAL is the only one that doesn't have included TV besides MRF-based regularization, and all improvements in reconstruction that it brings are caused by MRF-based regularization only (e.g. lattice driven hard-thresholding of image coefficients). Therefore the quality of reconstruction achieved by the LaSAL method serves us as

a lower bound during the development of the other three methods. In situations when we have a priori knowledge about the signal support or at least about image structure, which can help us to initialize MRF parameters, LaSAL should be our first choice. LaSAL2 is an improved version of LaSAL which needs more algorithm iterations than LaSAL to reach stable reconstruction and due to higher algorithm complexity (e.g. compound regularization), its expressiveness in signal representation can be reached with more measurements. Therefore, if we have, besides reasons for the choice of LaSAL, an opportunity to gather more measurements (e.g. multi-coil acquisition), LaSAL2 should be our choice. The other two algorithms CLaTV and FCLaTV differ from LaSAL2 in the usage of a more generic MRF signal support model (an anisotropic model), with automatic estimation of the MRF parameters, and a regularization based on soft-thresholding instead of hard-thresholding the image coefficients. A more general anisotropic MRF model can better represent signal support compared to LaSAL and LaSAL2 and with soft-thresholding instead of hard-thresholding, the reconstruction artifacts tend to be less pronounced. However, the increased expressiveness of the prior model together with a different update rule based on soft-thresholding comes at a price of a slower convergence of the CLaTV method. Therefore, we developed also its faster version of FCLaTV. Hence, if we can not be sure about the setup of MRF parameters or when the image under reconstruction can exhibit highly anisotropic structures, the algorithms derived within the FCSA framework i.e. CLaTV and FCLaTV are preferable choices. The presented results, obtained by thorough evaluation with different datasets acquired on the Cartesian and non-Cartesian grid with various sampling rates and trajectories, demonstrated significant improvement in MR image recovery compared to the other state-of-the-art methods which don't use MRF for the signal model. In the following chapter, we analyze the greedy based approach for MR reconstruction with an MRF signal model.

# 5

# Greedy reconstructions with MRF priors

*Above all, don't fear difficult moments. The best comes from them.*
—Rita Levi-Montalcini

In this chapter, we develop a greedy reconstruction method based on our proposed MRF prior. This method, as an alternative to the optimization-based methods from the previous chapter, is built on the foundation of greedy matching pursuit algorithms for signal recovery. The considered greedy approach reduces the algorithm parameters and the iteration steps compared to the optimization-based methods. We involve only MRF-based regularization that selects important image coefficients according to their estimated spatial support. We analyze how the resulting reduction of the overall complexity in the algorithm framework influences the reconstruction performances. We expect that the simplified greedy framework in combination with the adopted MRF model, which narrows the search space for the proper signal estimate, will lead to stable image recovery in a relatively small number of iterations. While enabling a considerable reduction of the computational complexity and the parameters, the developed greedy approach yields competitive results that are comparable with the results obtained from the optimization-based approaches. This also confirms that the advocated reconstruction framework with MRF priors yields consistent improvement in MR image reconstruction regardless of the choice of the iterative framework.

## 5.1 Related work

The greedy iterative methods represent the second group of CS recovery algorithms used to obtain the sparsest solution of the undetermined system in (3.22). These methods are less computationally complex and therefore much faster compared to the standard $\ell_1$ norm-based optimization techniques, but are also less precise. They are pursuing the sparsest signal representation in a chosen basis or frame $\mathbf{P}$ by selecting their elements, i.e., columns of $\mathbf{P}$ that participate in the creation of the given measurement vector $\mathbf{y}$ through the sensing operator $\mathbf{A}$. The well-known greedy algorithms which are commonly used in CS are listed in 3.4.

In the following, we will restrict our attention to the approach utilized in matching pursuit (MP) greedy methods. In subsection 3.4.6 we reviewed the optimization-based method LaSB which has strong similarities with the greedy Lattice Matching Pursuit (LaMP) method proposed in [Cevher 09b]. The LaMP algorithm, based on the extension of the MP approach named orthogonal matching pursuit (OMP), during reconstruction incorporates modeling of the spatial support of sparse images with a Markov Random Field (MRF). Since the most of MR images are compressible, i.e. approximately sparse in the transform domain, authors in [Pižurica 11] proposed the related algorithm LaSB, explained in subsection 3.4.6, which uses MRF to model the support of image coefficients and therefore considers the reconstruction of a broader class of MR images.

It was unclear so far whether the success of LaSB could also be reached with a simpler, greedy type of methods, and it was also not clear how any of these methods would compare to alternative wavelet-tree sparsity methods [Chen 12, Huang 12]. We address these questions and design a fast and simple greedy MRF-based method for CS-MRI which demonstrates excellent performance.

A preliminary version of this work has been reported as an abstract only, in [Panić 16b]. In the following we elaborate the method, explaining the details of the algorithm and we provide its thorough analysis and evaluation on real MRI images. This work complements optimization-based methods reported in the previous chapter. The method we are proposing here is conceptually much simpler, easier to implement and analyze compared to methods in the previous chapter, while it provides a similar improvement over the state-of-the-art wavelet-tree sparsity methods.

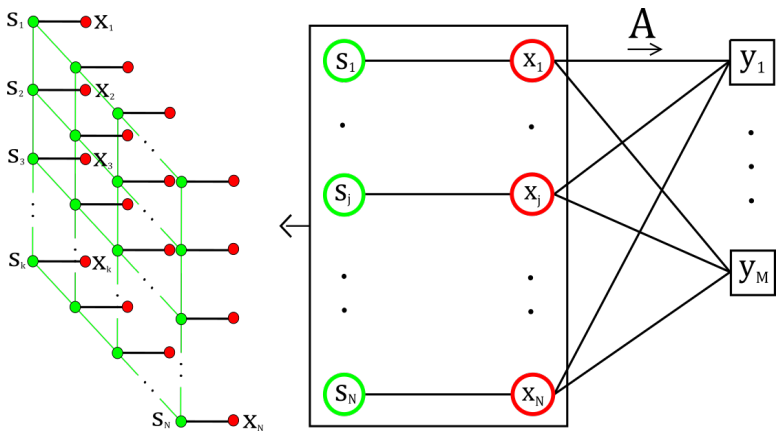



**Figure 5.1:** A graphical representation of reconstruction problem of sparse image $\mathbf{x}$ in canonical domain $\mathbf{P} = \mathbf{I}$ with involved MRF model $P(\mathbf{s})$ for the support.

## 5.2 Greedy Lattice pursuit method (GreeLa)

Let us first revisit briefly the original LaMP algorithm of [Cevher 09b], before analyzing possible extensions inside the greedy framework to make it applicable to MRI. Our new algorithm, inspired by this analysis, will follow then.

The original LaMP, with the pseudocode (using our notation) in Alg. 14, assumes that the image is sparse in the canonical (spatial) domain. Its main idea is to incorporate the estimation of the likely *support* $\mathbf{s}$ of the actual signal into the matching pursuit iterations. Since RIP property introduced in chapter 3 treats all possible $K$-sparse signal supports equally, incorporation of a probabilistic model on signal supports $\mathbf{s}$ and considering only the signal supports with the highest likelihoods, can potentially improve signal reconstruction in terms of the required number of measurements. In [Cevher 09b] authors proposed a graphical model for the CS acquisition process in (3.22) which represents joint probability distribution of signal $\mathbf{x}$, its support $\mathbf{s}$ and measurements $\mathbf{y}$ i.e. $p(\mathbf{s}, \mathbf{x}, \mathbf{y})$. With the following factorization $p(\mathbf{s}, \mathbf{x}, \mathbf{y}) = P(\mathbf{s})p(\mathbf{x}|\mathbf{s})p(\mathbf{y}|\mathbf{x})$ and by specifying the each factor, the solution $\mathbf{x}$ is obtained by maximizing the joint probability distribution. For support $\mathbf{s}$ they utilized a MRF prior or equivalently, according to the Hammersley-Clifford theorem [Li 09], a Gibbs distribution $P(\mathbf{s})$ given in equation (4.22). For an energy function $H(\mathbf{s})$ an Ising model defined on a rectangular lattice with

labels $s_i \in \{-1, 1\}$ is used defined in (4.23) and (4.24).[1]. Distribution of measurements $p(\mathbf{y}|\mathbf{x})$ follows the multivariate Gaussian distribution $\mathcal{N}(\mathbf{y}|\mathbf{A}\mathbf{x}, \sigma^2\mathbf{I})$ where $\sigma^2$ stands for the unknown noise variance. Signal likelihood $p(\mathbf{x}|\mathbf{s})$ is assumed to belong to the class of exponential distribution, which leads to the following form of joint distribution:

$$p(\mathbf{s}, \mathbf{x}, \mathbf{y}) \propto \exp\Big\{ \sum_{\langle i,j \rangle} \beta s_i s_j + \sum_i [\alpha s_i + \log(p(\mathbf{x}_i|s_i)] - \frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \Big\}. \tag{5.1}$$

In Alg. 14 we showed steps of LaMP algorithm which iteratively search for the estimate of the image $\mathbf{x}$ for which joint probability distribution in (5.1) has maximum value. Graphical representation of the $p(\mathbf{s}, \mathbf{x}, \mathbf{y})$ is depicted in Fig. 5.1. In particular, Step 4 in each iteration $k$ of Alg. 14 assigns to $\mathbf{s}^{\{k\}}$ the *maximum a posteriori* (MAP) estimate of the support of the temporary signal estimate $\mathbf{x}_t^{\{k\}}$, assuming a MRF prior $P(\mathbf{s})$ for the support. The temporary signal estimate $\mathbf{x}_t$ is obtained from the steps 1 and 2 using the gradient decent step obtained only considering $\ell_2$ norm term in (5.1). With a homogeneous Ising model and using the common conditional independence assumption for the likelihood $p(\mathbf{x}_t|\mathbf{s}) = \prod_i p([\mathbf{x}_t]_i|s_i)$, the MAP estimate of the support of $\mathbf{x}_t^{\{k\}}$ (denoted as *MAP-support*$\{\mathbf{x}_t^{\{k\}}\}$ in Alg. 14) is:

$$\mathbf{s}_{MAP}^{\{k\}} = \max_{\mathbf{s} \in [-1,1]^N} \sum_{\langle i,j \rangle} \beta s_i s_j + \sum_i [\alpha s_i + \log(p([\mathbf{x}_t^{\{k\}}]_i|s_i)]$$

The pseudo-inversion $\mathbf{A}^\dagger$ of the measurement matrix (Step 5) is then applied only for the columns of $\mathbf{A}$ selected by $\mathbf{s}^{\{k\}}$. Additional pruning to $K$ largest signal components (Step 6) yields the signal estimate $\mathbf{x}^{\{k\}}$. Although the objective function in (5.1) is nonconvex, a local optimum can be efficiently obtained via greedy iterations in Alg. 14 by alternating minimization over the support for the fixed signal and then over the signal for the estimated support.

This algorithm is directly applicable to the problem (3.23), only with $\mathbf{P} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix (see Fig. 5.1). We address possible algorithm extension such that it works in the case where $\mathbf{P}$ corresponds to a wavelet-like transform and hence for the reconstruction of compressible MR images.

[1] Although in [Cevher 09b], a *non-homogeneous model* is allowed in general with variable parameters $\beta_{i,j}$ and $\alpha_i$ depending on the spatial position, authors considered *homogeneous model* in experimental evaluation so we restrict our attention to this particular case.

**Algorithm 14** LaMP [Cevher 09b]

**Input:** $k = 1, \mathbf{y}, K, \mathbf{x}^{\{0\}}, \mathbf{t} = \mathbf{0}$

1: **repeat**{Matching Pursuit Iterations}
2: $\quad \mathbf{r}^{\{k\}} = \mathbf{y} - \mathbf{A}\mathbf{x}^{\{k-1\}}$
3: $\quad \mathbf{x}_t^{\{k\}} = \mathbf{A}^H \mathbf{r}^{\{k\}} + \mathbf{x}^{\{k-1\}}$
4: $\quad \mathbf{s}^{\{k\}} = \textit{MAP-support}\{\mathbf{x}_t^{\{k\}}\}$
5: $\quad \mathbf{t} = \mathbf{0}; \quad \mathbf{t}[\mathbf{s}^{\{k\}} = 1] = \mathbf{A}^{\dagger}[\mathbf{s}^{\{k\}} = 1, :]\mathbf{y};$
6: $\quad \mathbf{x}^{\{k\}} = \textit{Prune}(\mathbf{t}, K)$
7: $\quad k = k + 1$
8: **until** Maximum iterations or $\|\mathbf{r}^{\{\mathbf{k}\}}\| \leq \textit{threshold}$

A possible extension, which would allow applying LaMP to CS-MRI would be to replace steps 4-6 with:

$$\boldsymbol{\theta}_t^{\{k\}} = \mathbf{P}\mathbf{x}_t^{\{k\}}; \quad \mathbf{s}^{\{k\}} = \textit{MAP-support}\{\boldsymbol{\theta}_t^{\{k\}}\} \tag{5.2a}$$

$$\boldsymbol{\theta}_{t'}^{\{k\}} = \mathbf{P}\mathbf{A}^{\dagger}\mathbf{y}; \quad \mathbf{t}[\mathbf{s}^{\{k\}} = 1] = \boldsymbol{\theta}_{t'}^{\{k\}}[\mathbf{s}^{\{k\}} = 1] \tag{5.2b}$$

$$\boldsymbol{\theta}^{\{k\}} = \textit{Prune}(\mathbf{t}, K); \quad \mathbf{x}^{\{k\}} = \mathbf{P}^H \boldsymbol{\theta}^{\{k\}} \tag{5.2c}$$

Two important problems with this extension are: (i) the calculation of $\mathbf{P}\mathbf{A}^{\dagger}\mathbf{y}$ is costly, both in terms of the computation time and memory requirements and (ii) determining $K$ in each subband is not trivial. In [Cevher 09b] a few iterations of conjugate gradient (CG) method are used for efficiently performing the product by a pseudoinverse $\mathbf{A}^{\dagger}$ which additionally increased algorithm complexity. Hence, we propose a simplified, greedy algorithm where the computation of the pseudo-inverse (directly or with the CG) is avoided by replacing $\boldsymbol{\theta}_{t'}^{\{k\}}$ in (5.2b) by $\boldsymbol{\theta}_t^{\{k\}}$ and by excluding the additional pruning step (5.2c) (the sparseness is guaranteed already by the estimated support $\mathbf{s}^{\{k\}}$ using the right parameters of the prior MRF model). Unlike in [Pižurica 11] and the same as in the previous chapter we allow different a priori probabilities $\alpha \neq 0$ so that we can enforce the sparsity of the supports. The final image estimate is obtained by applying the hard-thresholding rule on image coefficients according to the estimated support $\mathbf{s}^{\{k\}}$, the same as for LaSAL and LaSAL2 method proposed in the previous chapter.

The proposed greedy algorithm named GreeLa (Greedy Lattice regularization) is summarized in Alg. 15. We employ the same likelihood and MRF model introduced in the previous chapter as well as the inference algorithm for the MAP estimation of the signal support. Although

**Algorithm 15** The proposed algorithm: GreeLa

**Input:** $k = 1, \mathbf{y}, \mathbf{x}^{\{0\}}, \mathbf{t} = \mathbf{0}$

**repeat**
  $\mathbf{r}^{\{k\}} = \mathbf{y} - \mathbf{A}\mathbf{x}^{\{k-1\}}$
  $\mathbf{x}_t^{\{k\}} = \mathbf{A}^H \mathbf{r}^{\{k\}} + \mathbf{x}^{\{k-1\}}$
  $\boldsymbol{\theta}_t^{\{k\}} = \mathbf{P}\mathbf{x}_t^{\{k\}}$
  $\mathbf{s}^{\{k\}} = \mathit{MAP\text{-}support}\{\boldsymbol{\theta}_t^{\{k\}}\}$
  $\mathbf{t} = \mathbf{0}; \quad \mathbf{t}[\mathbf{s}^{\{k\}} = 1] = \boldsymbol{\theta}_t^{\{k\}}[\mathbf{s}^{\{k\}} = 1]$
  $\boldsymbol{\theta}^{\{k\}} = \mathbf{t}, \mathbf{x}^{\{k\}} = \mathbf{P}^H \boldsymbol{\theta}^{\{k\}}$
  $k = k + 1$
**until** Maximum iterations or $\|\mathbf{r}^{\{\mathbf{k}\}}\| \leq \mathit{threshold}$

there is no theoretical guarantee for the convergence at this point, the proposed method converges in practice relatively fast.

### 5.2.1 Data sets and reference methods

We evaluate the proposed method on T1 MRI sagittal *brain slices* from the dataset introduced in subsection 4.5.4. Some of the slices are shown in Fig. 5.3. For measurements sub-sampling we simulate different trajectories: random lines, radial, Fibonacci spiral [Cline 01] and random sampling, illustrated in Fig. 5.2. All trajectories are defined as binary matrices on the Cartesian grid, which act as masks for selecting the corresponding Fourier coefficients. We provide results of comparison with the pFISTA method [Liu 16] on an image used in [Liu 16] shown in Fig. 5.3. The results are reported for simulated *radial* and *random* undersampling trajectories shown in Fig. 5.2.

We compare the results to LaSB [Pižurica 11], and to state-of-the art methods FCSA [Huang 11b], FCSANL [Huang 12] and WaTMRI [Chen 14] with their original implementations[2]. All these methods, except LaSB, employ a compound regularization. FCSA combines TV and $\ell_1$ norms while FCSANL combines non-local TV and $\ell_1$ norm. WaTMRI besides TV and $\ell_1$ norm involves overlapping groups in regularization as a approximation of tree-structured sparsity. In section 3.4 we provided more details about LaSB, WaTMRI and FCSA. We include results of image reconstruction of *pomelo* fruit which we used in experimental evaluation of methods proposed in previous chapter. The MRF parameters

[2] `http://ranger.uta.edu/~huang/index.html`

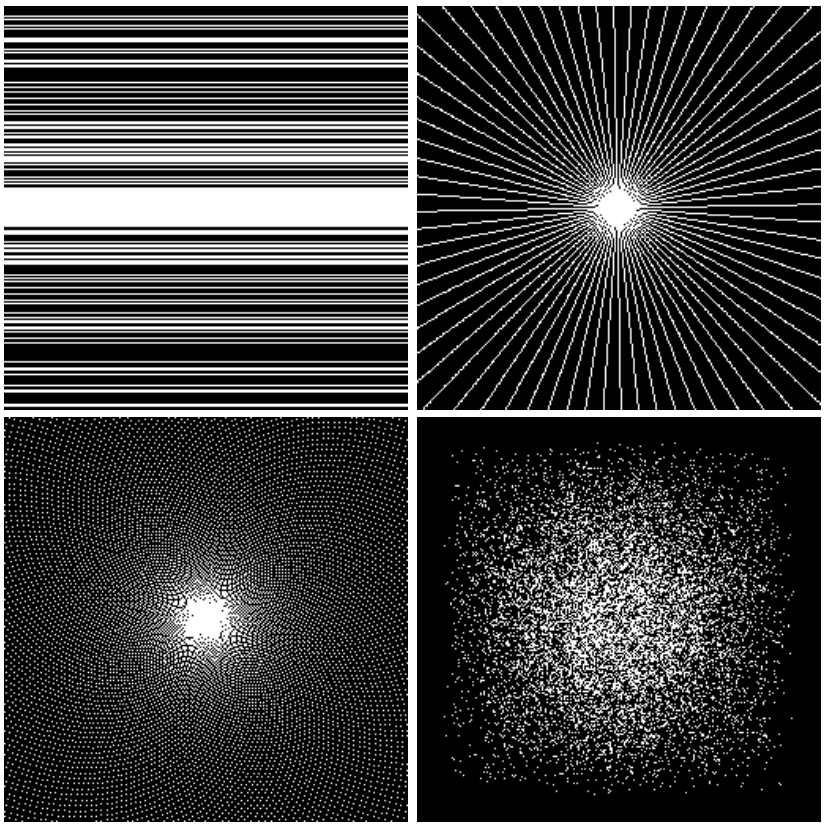


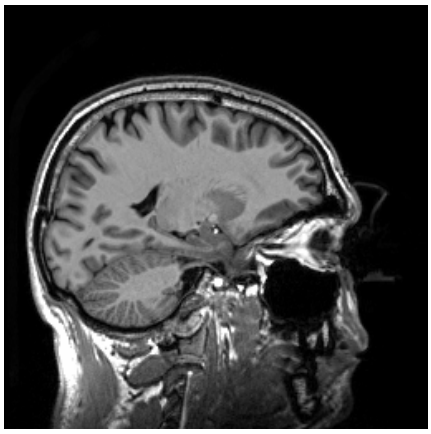
**Figure 5.2:** Examples of sampling trajectories used in the experiments: **Up** random lines, radial, **Bottom** Fibonacci spiral and random.

were optimized separately for LaSB ($\alpha = .017$, $\beta = .07$) and for GreeLa ($\alpha = 1e-4$, $\beta = .34$) and such as are used in all presented results.

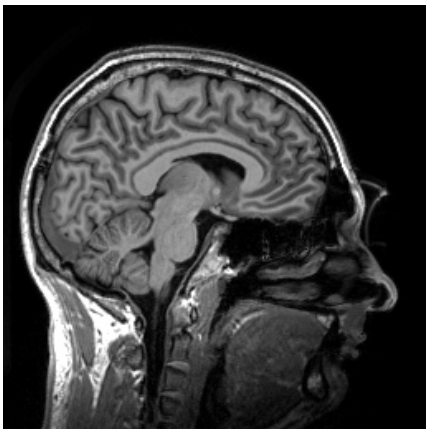
## 5.3 Experiments and Discussion

Here we report the results of extensive experiments on different MRI images, including an MRI data set and dataset of pomelo fruit that is directly acquired in $k$-space. We show the results on 248 sagittal slices from the T1 MRI dataset (each slice is a 256×256 image). For the sparsifying transform, we used the non-decimated wavelet transform with 3 scales and with 3 orientations per scale (fine-to-coarse) in all our experiments.

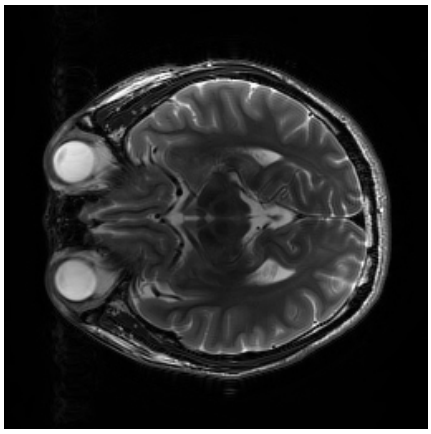
Fig. 5.4 shows the Peak Signal to Noise Ratio (PSNR) and



(a) *sagittal1*



(b) *sagittal2*



(c) *axial1*

**Figure 5.3:** Test images *sagittal1* and *sagittal2* are two slices from data set comprising 248 images. The bottom image *axial1* is from [Liu 16]. All images are 256×256.

Structural Similarity Index (SSIM) for one slice (the first image in Fig. 5.3), with sampling rate (SR) ranging from 14% to 48%, and the evolution of the PSNR and SSIM per iteration for a particular SR (20%). The MRF-based methods GreeLa and LaSB achieve a consistent and significant improvement in PSNR (more than 4 dB) compared to WaTMRI, FCSA, and FCSANL for all SR values, and they also approach convergence in fewer iterations. GreeLa yields slightly higher PSNR than LaSB and shows a more stable behavior in the first 20 iterations (see bottom

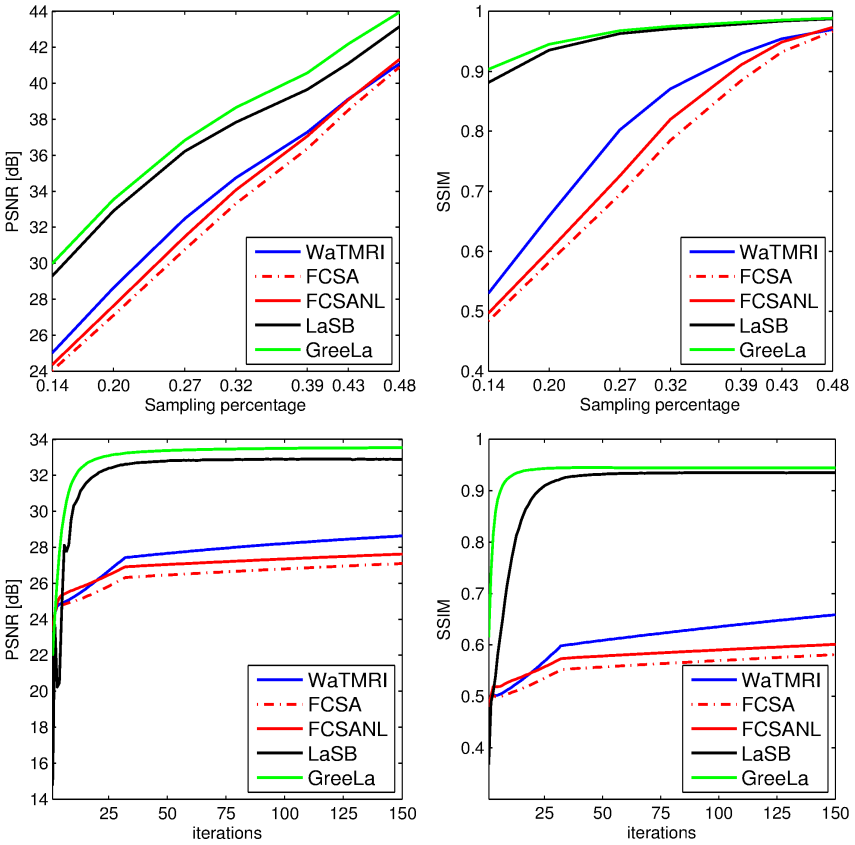



**Figure 5.4: Top left** and **Top right**: PSNR and SSIM for the reconstructions of one slice (the first in Fig. 5.3) at different sampling rates. **Bottom left** and **Bottom right**: Reconstruction performances in PSNR and SSIM, respectively on the same slice with 20% measurements in 150 iterations.

left in Fig. 5.4). In the case of SSIM measure LaSB and GreeLa outperform compared methods significantly for all sampling rates (see the top right diagram in Fig. 5.4). LaSB and GreeLa reached SSIM above 0.85 for all SR, GreeLa even more than 0.9 for an SR of 14%. For SR of 20%, LaSB and GreeLa reached the SSIM above 0.9 in less than 20 iterations (see bottom right in Fig. 5.4) while among the compared methods WaTMRI performed best with SSIM above 0.65 after 150 iterations. This significant structural difference in reconstruction for a low SR is presented in Fig. 5.5.

We show results of reconstruction of all 248 MRI sagittal slices from our dataset in Fig. 5.6 with SR=48%. Here we show the only

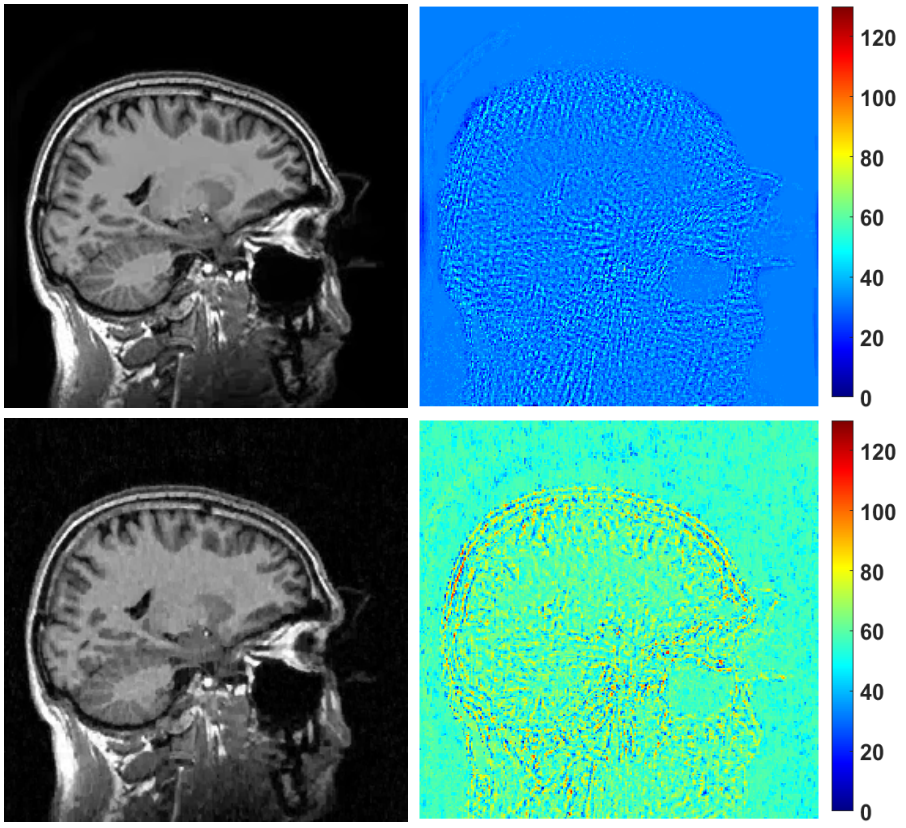



**Figure 5.5:** Reconstructed image (the first in Fig. 5.3) from 20% of measurements using radial trajectory. **Top** GreeLa and **Bottom** WaTMRI algorithm. The images on the right show reconstruction errors.

comparison with WaTMRI since it outperforms FCSA and FCSANL on slices from this data set (see Fig. 5.4). The conclusions are as follows: although WaTMRI increased its performance on average, GreeLa and LaSB yield a superior PSNR and converge in fewer iterations. A more stable behavior of GreeLa compared to LaSB and slightly better PSNR are again observed.

We next compared GreeLa with pFISTA [Liu 16] using the image from [Liu 16] (see Fig. 5.3). We now use a random and radial sampling trajectory with a sampling rate of 30%. From the left diagram in Fig. 5.7 for the case of radial sampling trajectory, GreeLa reaches only slightly higher PSNR (35.3 dB) compared to the best version of pFISTA (35.1 dB). However, in the case of random sampling (the right-side dia-

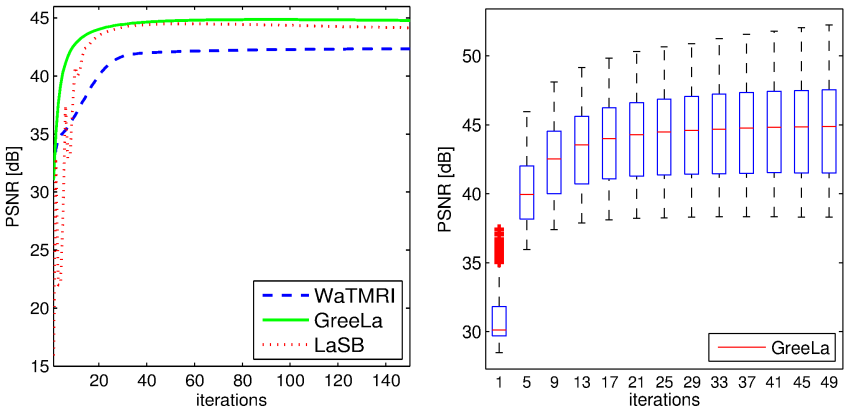



**Figure 5.6:** PSNR values obtained from 248 MRI *brain slices* from the first data set, with radial sampling. Mean PSNR (**Left**) and the PSNR distribution for GreeLa (**Right**). The results are presented as a box plot: the edges of each box represent $25^{th}$ and $75^{th}$ percentile while the central mark (red line) in the box is median. The whiskers extend to the most extreme PSNR values which are not considered outliers while outliers are plotted separately with red crosses.

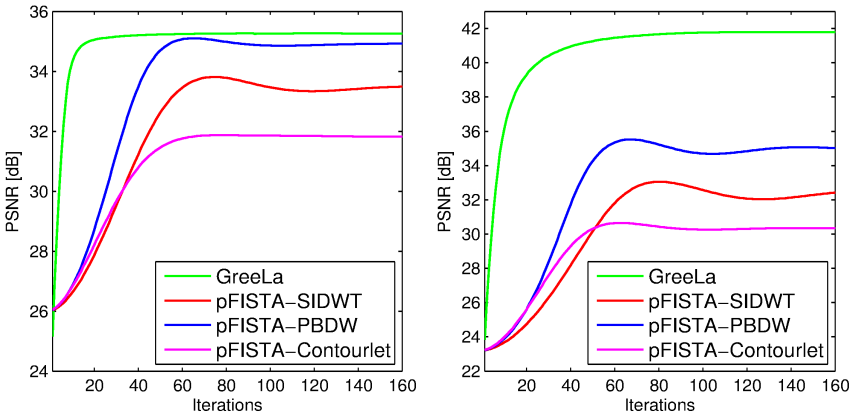



**Figure 5.7:** PSNR for the reconstructions of the *axial1* test image in Fig. 5.3 for different sampling trajectories. **Left**: radial and **Right**: random with the same sampling rate of 30%.

gram in Fig. 5.7), GreeLa yields a huge improvement of more than 6 dB compared to the best performing pFISTA variant.

Next, we perform experiments on a real MRI dataset obtained by scanning the pomelo fruit. The experimental setup corresponds to

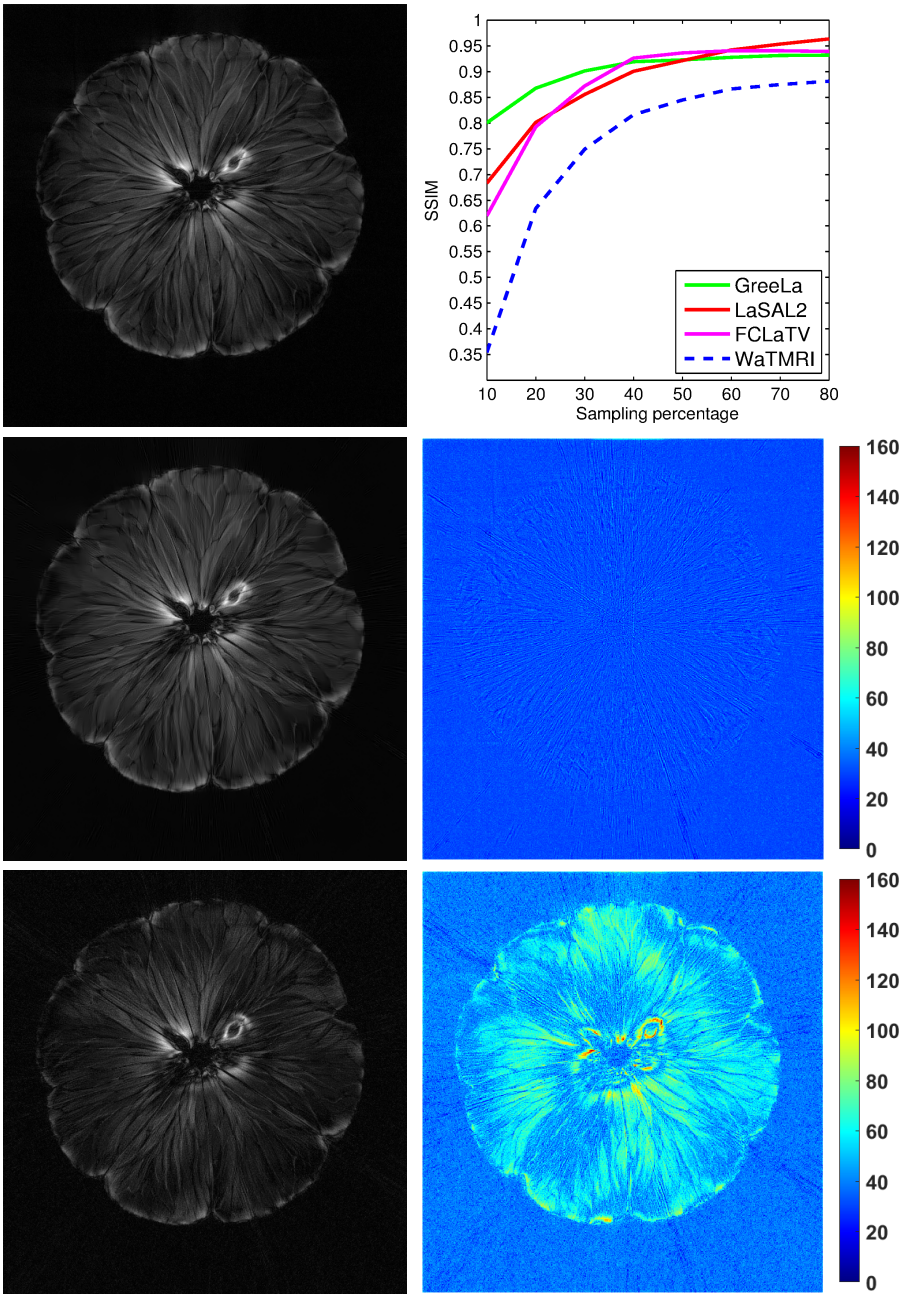



**Figure 5.8:** Pomelo experiment. **The first column top to bottom**: reference image obtained from 100% of measurements, reconstructions from 20 % sampling rate using GreeLa and WaTMRI respectively. **The second column top to bottom**: Obtained SSIM for different sampling rates and different reconstruction methods, followed by properly scaled error according to the corresponding reconstructions obtained by GreeLa and WaTMRI respectively.

the one used in testing the LaSAL and LaSAL2, which includes undersampling with golden ratio profile spacing [Winkelmann 07], weighting the non-uniform measurements based on an area of Voronoi cells and involving of non-uniform FFT procedures [Fessler 03]. The three reference methods (WaTMRI, FCSA, and FCSANL) give similar results on this image, so we choose for comparison WaTMRI. We also provide performances of LaSAL2 and FCLaTV on the used experimental setup. Fig. 5.8 shows visual comparison and SSIM values for GreeLa, LaSAL2, FCLaTV and WaTMRI. For all sampling rates, the proposed method GreeLa outperforms WaTMRI. GreeLa outperforms LaSAL2 and FCLaTV for sampling rates below 40%. In comparison with LaSAL2 only, GreeLa has better or similar performance up to a sampling rate of 50%. Slightly better SSIM measures achieved by FCLaTV and LaSAL2 for higher sampling rates are induced by a more complex iterative reconstruction framework which involves besides MRF a TV regularization as well. With more measurement included in reconstruction, subtle structures in the image can be revealed by FCaLaTV and LaSAL2, while GreeLa stays limited in this regard with the greedy framework and only MRF based regularization. However, given that the new algorithm is conceptually simpler, easier to implement and optimize and without involving TV regularization, these results are highly encouraging.

## 5.4 Conclusion

This chapter was dedicated to the development of a greedy method for MRI reconstruction which utilized the proposed MRF based regularization introduced in the previous chapter. Based on a LaMP iterative framework, developed for reconstruction of images sparse in the spatial domain, we proposed a related GreeLa algorithm for recovering of compressible MRI images. The simplified iterative procedure from the LaMP method is kept in GreeLa while the algorithm complexity is contained in signal support estimation now in the transform domain instead of in the spatial domain like it was in LaMP. The sparse solution is forced by hard thresholding rule on image coefficients according to the estimated support. In experiment evaluation, GreeLa outperforms state-of-the-art methods such as pFISTA, WaTMRI, FCSA, and FCSANL and showed comparable performances, for some sampling rates even better, as LaSB, LaSAL2, and FCLaTV.

# 6

# Multi coil MRI reconstruction

*The truth is rarely pure and never simple.*
—Oscar Wilde

Improvements in reconstructing Magnetic Resonance Images (MRI) from partial data have been reported in the previous two chapters using spatial context modeling with Markov Random Field (MRF) priors. However, these algorithms have been developed only for magnitude images from single-coil measurements. In practice, most of the MRI images today are acquired using multi-coil data. In this chapter, we extend our approach for MRI reconstruction with MRF priors to deal with multi-coil data i.e., to be applicable in parallel MRI (pMRI) settings.

## 6.1 Introduction to multi-coil image reconstruction

Slow acquisition of magnetic resonance images (MRI) is still a barrier in everyday clinical usage. Parallel imaging in MRI (pMRI) offers a significant speedup by acquiring simultaneously signals from multiple coils [Blaimer 04]. In subsections 2.1.3 and 2.2.3 we introduced pMRI reconstruction principles and reviewed some of the commonly used methods. Another approach for speeding up the acquisition is applying clever algorithms for image reconstruction from partial data reviewed in chapter 3, which are often termed as compressed sensing in MRI (CS-MRI) [Daubechies 04, Donoho 06a]. The potential of compressed sensing has already been demonstrated in pMRI [Jin 16, Chun 15, Aelterman 10, Liang 09, Liu 08]. In the following, we extend our CS-MRI

method LaSAL2 from chapter 4 to the pMRI setting.

For the sake of clarity and completeness recall the mathematical problem formulation from chapter 2 for the single-coil MR acquisition, where the estimation of image $\mathbf{x}$ is obtained as a solution of the following optimization problem

$$\min_{\mathbf{x}\in\mathbb{C}^N} \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2^2 + \tau\phi(\mathbf{x}) \tag{6.1}$$

where $\frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2^2$ is a data fidelity term, $\phi : \mathbb{C}^N \mapsto \mathbb{R}\cup\{-\infty,+\infty\}$ is a regularization function and the parameter $\tau$ denotes the amount of the involved regularization. The pMRI reconstruction problem can be formulated in image domain (SENSE method [Pruessmann 99]) or in $k$-space domain (GRAPPA method [Blaimer 04]). In this work, we opt for the maximum likelihood (ML) formulation of the multi-coil reconstruction problem given as

$$\min_{\mathbf{x}\in\mathbb{C}^N} \sum_{i=1}^{N_c} \|\mathbf{y}_i - \mathbf{F}\mathbf{C}_i\mathbf{x}\|_2^2 \tag{6.2}$$

where matrix $\mathbf{F}$ represents the Fourier transform, the index $i$ refers to the $i$-th coil, with coil data $\mathbf{y}_i$ and the coressponding spatial sensitivity $\mathbf{C}_i$. In the case of a single coil ($i=1$), equation (6.2) reduces to a form similar to the data fidelity term in (6.1). The difference is that instead of the operator $\mathbf{A}$, we now have $\mathbf{F}\mathbf{C}_i$, where different k-space components are attenuated according to the particular coil sensitivities, but there is no strict undersampling. Since we are dealing with undersampled measurements from coils, which are in sum still not sufficient to have a well-posed reconstruction problem, some sort of regularization needs to be involved. In the next section, we extend our LaSAL2 method for single coil reconstruction from Section 4.4 to the more general pMRI-CS problem.

## 6.2 Related work on parallel imaging in MRI

In the pMRI reconstruction, the multiple $k$-space data $\mathbf{y}_i$ obtained from different receiver antenna coils are utilized in order to recover the MR image. The sensitivity encoding for fast MRI (SENSE) method from [Pruessmann 99] combines the aliased coil images $\mathbf{x}_i$, reconstructed using the inverse discrete two-dimensional Fourier transform (2D-IDFT) from each coil measurements $\mathbf{y}_i$ (or density corrected adjoint non-uniform

Fourier transform in the case of non-Cartesian $\mathscr{k}$-space trajectories), and creates a composite MR image with the a priori knowledge of coils sensitivities. Although an important issue with SENSE is the estimation of the coil sensitivity profiles, it is to date the most widely employed pMRI technique offered by many companies. For more details about the SENSE method, the reader is directed to chapter 2. The issue with the estimation of the sensitivity profiles can be circumvented with the generalized autocalibrating partially parallel acquisitions (GRAPPA) method from [Blaimer 04] which doesn't require knowledge of the coil sensitivities. It reconstructs the missing $\mathscr{k}$-space data using its adjacent neighborhood in $\mathscr{k}$-space from all coils. GRAPPA uses a kernel that defines how the missing $\mathscr{k}$-space samples from the $i^{th}$ coil data are interpolated using the acquired $\mathscr{k}$-space data from all of the coils. The interpolation process is learned from a fully sampled low-frequency part of the spectrum of every coil (so-called auto-calibration signal (ACS) [Griswold 02]). In chapter 2 we provide more information about GRAPPA method. After interpolating the missing $\mathscr{k}$-space data, a 2D-IDFT is employed to obtained coil images $\mathbf{x}_i$. The composite MR image $\mathbf{x}$ is usually reconstructed with the Sum of Squares (SoS) technique:

$$\mathbf{x} = \sqrt{\sum_{i=1}^{N_c} |\mathbf{x}_i|^2}. \tag{6.3}$$

Another common approach for pMRI reconstruction called iterative self-consistent parallel imaging reconstruction (SPIRiT) [Lustig 10] reconstructs each coil image separately and then combines the independent reconstructions into the composite MR image. Missing $\mathscr{k}$-space points are estimated from the acquired and reconstructed missing points in their neighborhood from all the coils. SPIRiT achieved better noise reduction and more accurate reconstruction compared to traditional GRAPPA-like approaches. A method referred to as ESPIRiT [Lai 10] bridges the gap between SENSE and GRAPPA and uses eigenvalue decomposition in image space to compute more robust sensitivity maps. It combines advantages of SPIRiT and GRAPPA methods and restricts the solution to a subspace spanned by the coil sensitivity maps.

When the data sampling density is below the Nyquist rate, the reconstruction problem becomes ill-posed and requires regularization. A compressive parallel sensing method COMPASS [Aelterman 10] expresses a multi-coil reconstruction problem as a basis pursuit (BP) optimization problem with $\ell_1$ norm as regularization on image sparse representation. An alternative method, termed LORAKS [Haldar 16,Kim 17]

employes low-rank modeling of local $\mathscr{k}$-space neighborhoods for parallel imaging with partial Fourier acquisition. It takes into account the fact that many MRI images have limited spatial support and smoothly varying phase through the constraints in the optimization framework.

In the following, we propose a new method for the pMRI-CS reconstruction problem, based on a recent approach for MRI reconstruction, with Markov random field priors [Panić 17a]. We generalize this method such to treat the data from the multiple coils simultaneously in a similar manner as in the COMPASS method from [Aelterman 10], but incorporating also a spatial context model.

## 6.3 pMRI-CS with MRF priors

We start by formulating the pMRI-CS problem as follows:

$$\min_{\mathbf{x}\in\mathbb{C}^N} \phi(\mathbf{P}\mathbf{x}) \quad \text{subject to} \quad \sum_{i=1}^{N_c} \|\mathbf{y}_i - \mathbf{A}\mathbf{C}_i\mathbf{x}\|_2^2 \leq \epsilon \tag{6.4}$$

where $\mathbf{P}$ models the linear sparsifying transform, $\epsilon \geq 0$ denotes a parameter which is related to the noise variance and other variables are as defined earlier. Since the measurements from the coils $\mathbf{y}_i$ are undersampled with the operator $\mathbf{A}$, some regularization is necessary in order to find a suitable solution. We introduce the following notation for the augmented measurement vector composed of multiple coil data $\mathbf{y}_a = [\mathbf{y}_1^T, \mathbf{y}_2^T ... \mathbf{y}_{N_c}^T]^T$ and the corresponding augmented vectorized image $\mathbf{x}_a = \mathbf{T}\mathbf{x}$, where $\mathbf{T}$,

$$\mathbf{T} = \begin{bmatrix} \mathbf{I}_{(1)}, & \mathbf{I}_{(2)}, & \mathbf{I}_{(3)}, & \dots & \mathbf{I}_{(N_c)} \end{bmatrix} \tag{6.5}$$

is formed by stacking row-wise the identity matrix $\mathbf{I}_{N\times N}$, $N_c$ times. With the involved augmented variables we can reformulate the problem from (6.4) as

$$\min_{\mathbf{x}\in\mathbb{C}^N} \phi(\mathbf{P}\mathbf{x}) \quad \text{subject to} \quad \|\mathbf{y}_a - \mathbf{A}_B\mathbf{C}_B\mathbf{x}_a\|_2^2 \leq \epsilon \tag{6.6}$$

where $\mathbf{A}_B$ is a block diagonal matrix, composed of the repeated matrices $\mathbf{A}$ along the diagonal while the block diagonal matrix $\mathbf{C}_B$ is formed by placing the coil sensitivity maps $\mathbf{C}_i$ along the diagonal:

$$\mathbf{A}_B = \begin{bmatrix} \mathbf{A}_{(1)} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{A}_{(N_c)} \end{bmatrix}, \mathbf{C}_B = \begin{bmatrix} \mathbf{C}_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{C}_{N_c}. \end{bmatrix} \tag{6.7}$$

**Algorithm 16** The proposed MRF-based reconstruction method for partially sampled pMRI.

**Input:** $k=0, n=10, \mu_1, \mu_2 > 0, \mathbf{v}^{\{0\}}, \mathbf{w}^{\{0\}}, \mathbf{z}^{\{0\}}, \mathbf{b}^{\{0\}}, \mathbf{c}^{\{0\}}, \mathbf{d}^{\{0\}},$

**repeat**
- $\mathbf{r}^{\{k\}} = \mu_1(\mathbf{z}^{\{k\}} + \mathbf{c}^{\{k\}}) + \mathbf{T}^H\mathbf{C}_B^H\mathbf{A}_B^H(\mathbf{v}^{\{k\}} + \mathbf{b}^{\{k\}})$
- $\mathbf{x}^{\{k+1\}} = \mathbf{ConjGrad}(\mathbf{r}^{\{k\}}, n)$
- $\mathbf{v}^{\{k+1\}} = \boldsymbol{\Psi}_{\iota_{E(\epsilon,\mathbf{I}\mathbf{y}_a)}}(\mathbf{A}_B\mathbf{C}_B\mathbf{T}\mathbf{x}^{\{k+1\}} - \mathbf{b}^{\{k\}})$
- $\mathbf{z}' = \frac{1}{(\mu_1+\mu_2)}\big(\mu_1(\mathbf{x}^{\{k+1\}} - \mathbf{c}^{\{k\}}) + \mu_2(\mathbf{w}^{\{k\}} + \mathbf{d}^{\{k\}})\big)$
- $\mathbf{z}^{\{k+1\}} = \boldsymbol{\Psi}_{\mathrm{TV}}(\mathbf{z}'; \mu_1 + \mu_2)$
- $\boldsymbol{\theta}' = \mathbf{P}(\mathbf{z}^{\{k+1\}} - \mathbf{d}^{\{k\}})$
- $\hat{\mathbf{s}} \leftarrow \mathit{MAP\text{-}support}\{\boldsymbol{\theta}'\}$
- $\mathbf{w}^{\{k+1\}} = \mathbf{P}^H(\boldsymbol{\theta}' \circ \hat{\mathbf{s}})$
- $\mathbf{b}^{\{k+1\}} = \mathbf{b}^{\{k\}} - (\mathbf{A}_B\mathbf{C}_B\mathbf{T}\mathbf{x}^{\{k+1\}} - \mathbf{v}^{\{k+1\}})$
- $\mathbf{d}^{\{k+1\}} = \mathbf{d}^{\{k\}} - (\mathbf{z}^{\{k+1\}} - \mathbf{w}^{\{k+1\}})$
- $\mathbf{c}^{\{k+1\}} = \mathbf{c}^{\{k\}} - (\mathbf{x}^{\{k+1\}} - \mathbf{z}^{\{k+1\}})$
- $k = k+1$

**until** some stopping criterion is satisfied

In [Ramos-Llordén 16] authors used the similar variable transformation and augmentation in order to extend their method to multi-coil scenario. The constrained optimization problem in (6.6) can be transformed to its unconstrained form by adding to $\phi(\mathbf{P}\mathbf{x})$ an indicator of the feasible set $E(\epsilon, \mathbf{I}, \mathbf{y}) = \{\mathbf{x} \in \mathbb{C}^N : \|\mathbf{A}_B\mathbf{C}_B\mathbf{x}_a - \mathbf{y}_a\|_2 \leq \epsilon\}$ as follows:

$$\min_{\mathbf{x}\in\mathbb{C}^N} \phi(\mathbf{P}\mathbf{x}) + \boldsymbol{\iota}_{E(\epsilon,\mathbf{I},\mathbf{y}_a)}(\mathbf{A}_B\mathbf{C}_B\mathbf{x}_a). \tag{6.8}$$

A solution of the obtained unconstrained optimization problem in (6.6) at the same time must satisfy two conditions. First, the expected signal structure imposed by the regularization function must be obtained. Second, a solution must be close enough in the $\mathcal{k}$-space to measurements from all coils, where proximity between the measurements and solution is defined through the feasible set condition. For the regularization function in (6.8) we opt for a linear combination of the total variation (TV) and MRF regularization functions, which already showed good performances in the LaSAL2 from [Panić 17a] for the single-coil reconstruction problem. Following the same derivation as for the LaSAL2 method in chapter 4 we solve the resulting unconstrained problem (6.8) with a compound regularization function (TV + MRF) by the alternating direction method of multipliers (ADMM). The resulting

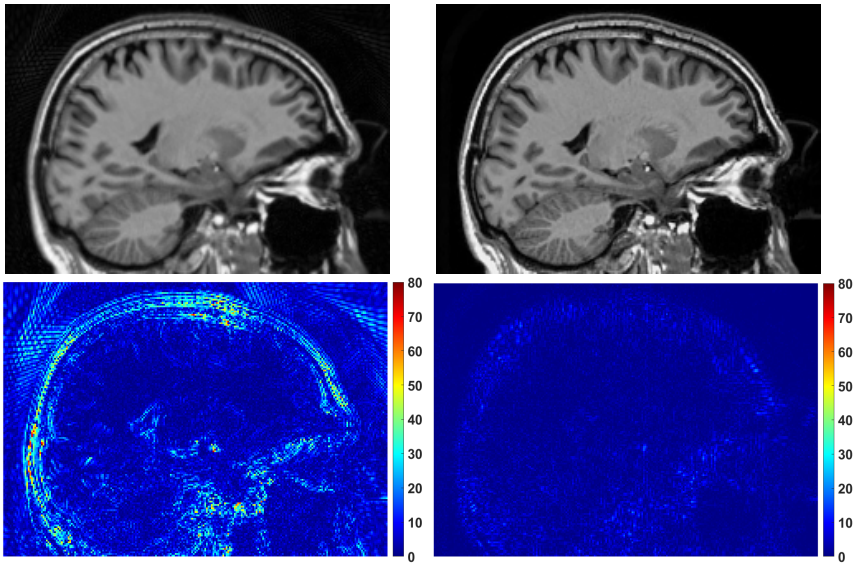



**Figure 6.1:** Reconstruction of a sagittal slice from 4 coils with 12% of measurements per coil sampled with spiral trajectory. **First row**: Images reconstructed with COMPASS [Aelterman 10] (27.75 dB) and the proposed method (35.42 dB), respectively. **Second row**: The corresponding reconstruction errors.

method is summarized under Algortihm 16. The derived method differs from LaSAL2 in two aspects: utilization of multiple coil measurements and usage of conjugate gradient (CG) iterations in step 3. Instead of using the SMW matrix inversion lemma [Deng 11](see 3.4.3), a temporary image estimate is obtained with relatively few iterations (up to 10) of a CG method. This small change in algorithm complexity compared to the LaSAL2 method is reflected in a negligible increase of time needed for one algorithm iteration.

## 6.4 Experiments and discussions

In the first experiment we compared our proposed method with the method from [Aelterman 10] method on a sagittal MRI slice acquired using measurements from 4 coils sampled equiangularly along a spiral $\mathcal{k}$-space trajectory with an average sampling density that corresponds to 12% of the Nyquist rate. Despite using pMRI to quadruple the number of data samples, this still corresponds to an ill-posed CS-pMRI problem since the overall number of measurements is less than the Nyquist

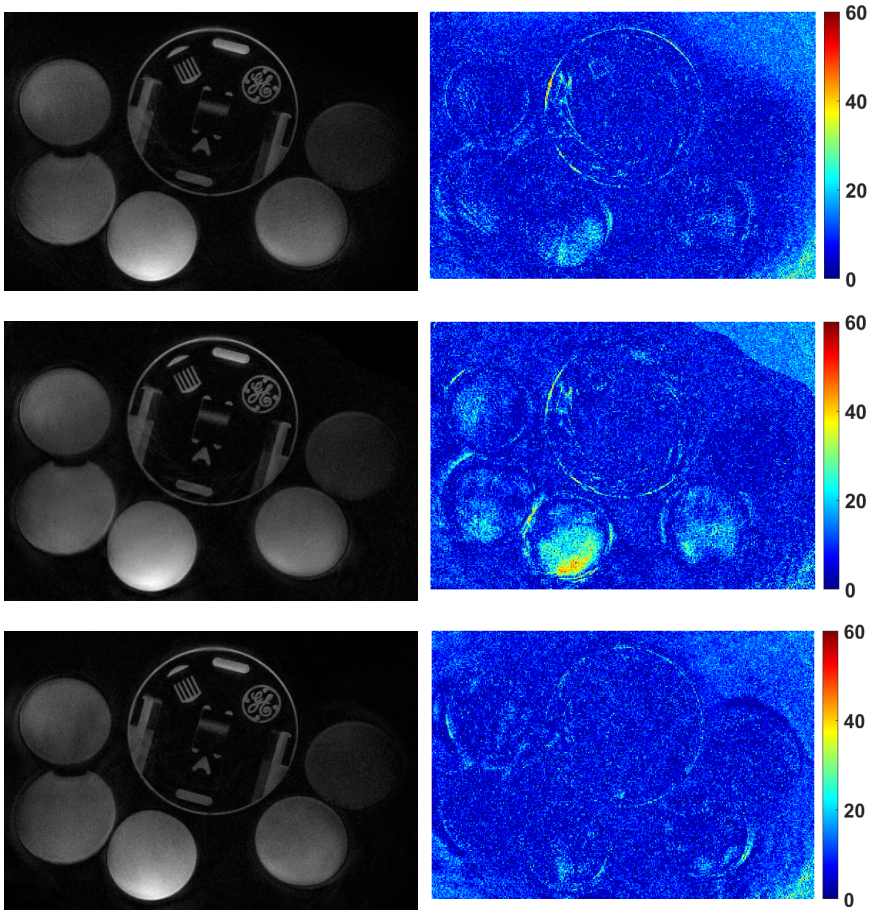



**Figure 6.2:** Reconstructions of a phantom test image using measurements sampled with the acceleration factor 3 from 4 virtual coils. **First column**: Reconstructed images with SPIRIT (32.22 dB), ESPIRIT (31.32 dB) and the proposed method (32.64 dB) respectively. **Second column**: The corresponding reconstruction errors.

sampling rate. The results, shown in Fig. 6.1, demonstrate the advantage of our regularization based on MRF prior compared to the $\ell_1$ norm regularization in the multi-coil reconstruction scenario. Although both methods used the same coil sensitivity maps, the proposed approach yields an improvement in the peak signal to noise ratio (PSNR) of more than 6dB. The reconstructed image is much sharper and contains much more original details, as it also evident from the difference image.

The next experiment was evaluated on publicly available data

| Methods | SSIM | NRMSE |
|---|---|---|
| P-LORAKS | 0.88 / 0.92 | 0.07 / 0.06 |
| SENSE-LORAKS | 0.83 / 0.87 | 0.08 / 0.07 |
| Proposed method | **0.95** / **0.96** | **0.04** / **0.03** |

**Table 6.1:** Reconstruction of the T1 weighted brain image from [Kim 17] using undersampled measurements obtained from 4 coils with a uniform / random trajectory. A comparison of P-LORAKS and SENSE-LORAKS methods with the proposed one is given in terms of structural similarity measure (SSIM) and normalized root-mean-squared-error (NRMSE).

from `http://www.eecs.berkeley.edu/~mlustig` which is used in the experimental evaluation of ESPIRIT [Uecker 14] and SPIRIT [Lustig 10]. The lower sampling rate is obtained by decimating the data with an acceleration factor greater than 1. We use measurements undersampled with the acceleration factor 3 from 4 virtual coils using the SVD coil compression technique applied on 8 coils. This experimental setup is the same as the one used in [Uecker 14, Lustig 10]. In the proposed method the coil sensitivity maps are estimated from undersampled measurements using the approach from [Aelterman 14b]. From the presented results in Fig. 6.2 we see that the proposed method outperforms all the compared methods. Image artifacts caused by undersampling are better suppressed by the proposed method than by any of the compared methods.

We next compare our method with the P-LORAKS and SENSE-LORAKS methods from [Haldar 16] and [Kim 17]. We use the four-channel T1 weighted brain image from [Kim 17] with provided coil sensitivity profiles. Reconstruction was performed from only 14% of the measurements from each coil sampled using two different trajectories: random and uniform. A golden standard image is created using all measurements from four coils with the SoS technique. The obtained reconstruction performances are given in Table. 6.1. The proposed method yielded the best results in terms of both structural similarities and mean squared error. Fig. 6.3 and Fig. 6.4 show visual results for the uniform and random sampling trajectory, respectively. It can be seen that the image reconstructed with the proposed method contains much less noise compared to the reconstructions obtained by P-LORAKS and SENSE-LORAKS.

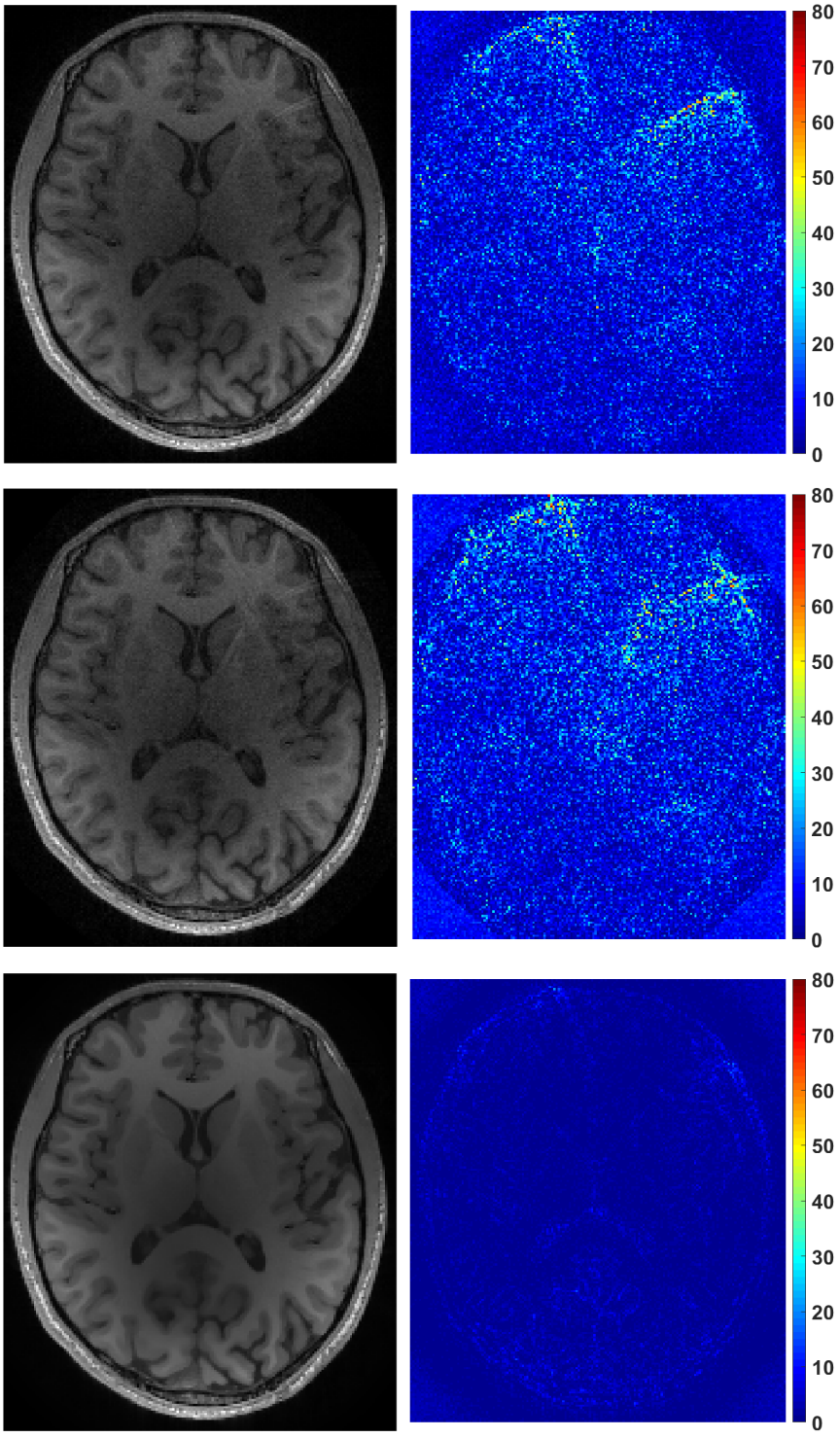



**Figure 6.3:** Reconstructions of the T1 weighted brain image used in [Kim 17] form uniformly sampled measurements from 4 coils. **First column**: Reconstructed images with P-LORAKS, SENSE-LORAKS and the proposed method respectively. **Second column**: The corresponding reconstruction errors.

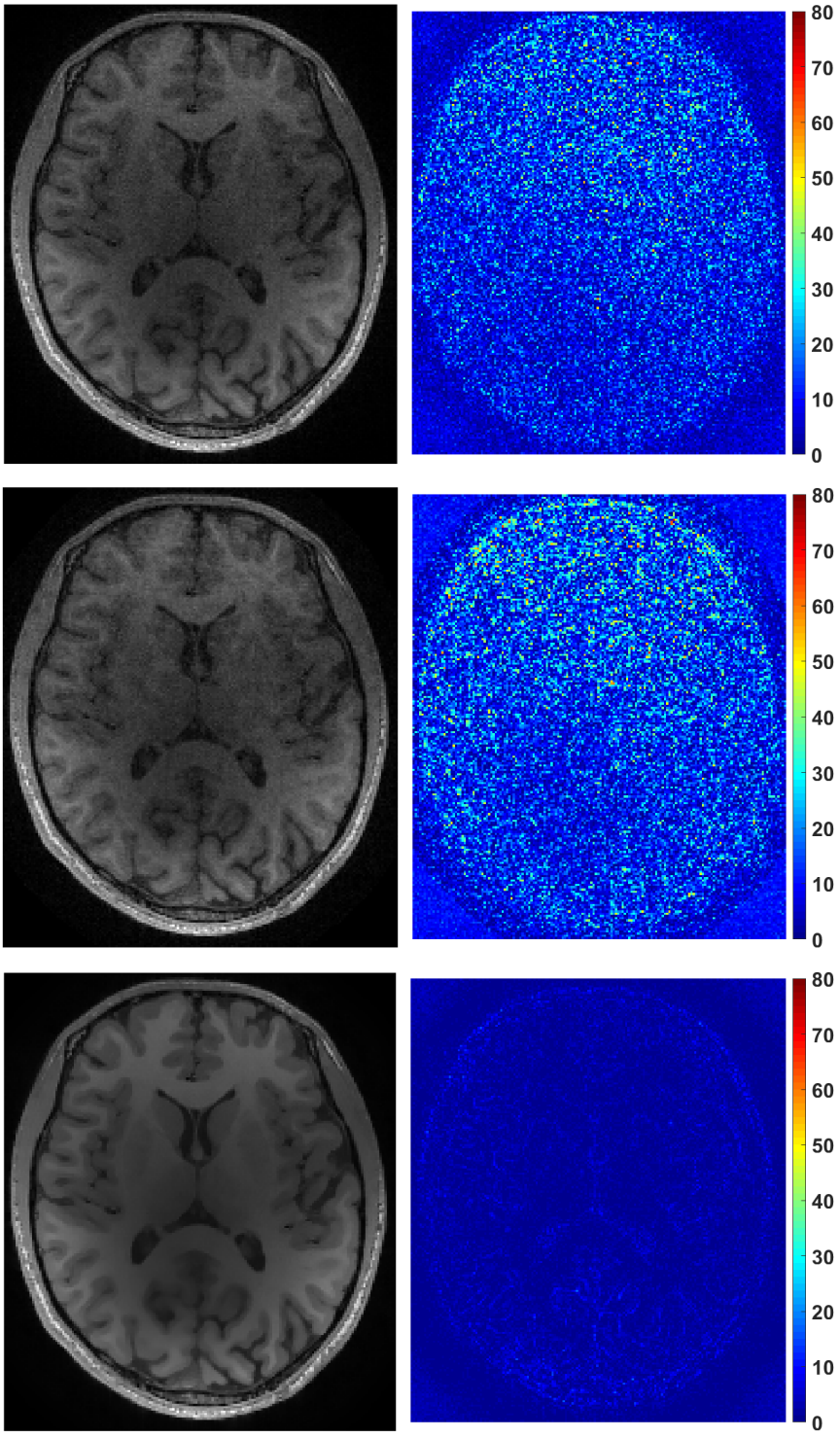



**Figure 6.4:** Reconstructions of the T1 weighted brain image used in [Kim 17] form randomly sampled measurements from 4 coils. **First column**: Reconstructed images with P-LORAKS, SENSE-LORAKS and the proposed method respectively. **Second column**: The corresponding reconstruction errors.

## 6.5 Conclusion

In this chapter we extend our LaSAL2 method from Chapter 4 to handle the multi-coil reconstruction. We considered undersampled measurements from all available coils i.e. pMRI-CS reconstruction problem. The proposed compound MRF + TV regularization is applied to a temporary composite MRI image estimate obtained after a few iterations of the CG method during the reconstruction procedure. The derived approach is compared with state of the art methods: COMPASS [Aelterman 10], SPIRIT [Lustig 10], ESPIRIT [Uecker 14], P-LORAKS [Haldar 16] and SENSE-LORAKS [Kim 17]. On various datasets, our proposed method significantly improved reconstruction performances compared to other tested methods. Visually, signal structure in our reconstructed image is better preserved and artifacts, due to undersampled measurements, are more suppressed in comparison with the reconstructed image from recently published methods P-LORAKS and SENSE-LORAKS.

# 7

# Conclusion

*The cleverest of all, in my opinion, is the man who calls himself a fool at least once a month.*
—Fyodor Dostoyevsky

In this thesis, we addressed the problem of magnetic resonance imaging (MRI) reconstruction from partial measurements using the compressed sensing (CS) theory and Markov random field (MRF) model for the image support in wavelet-like domains. The proposed MRF model reduced the space of possible solutions during reconstruction by enforcing the expected image structure under wavelet-like transformations. Image coefficients in transform domain are often clustered around important image structures, such as edges and textures, and are well represented through MRF-based spatial context modeling which we exploited in the proposition of reconstruction methods in this thesis.

Our original contribution is in the development of reconstruction methods, greedy and optimization-based, which utilized MRF based signal prior. We start with an optimization-based framework for the recovery of compressively sampled MRI (CS-MRI) data. In Chapter 4 we presented a comprehensive study and developed a novel algorithm that incorporates the MRF modeling framework into a constrained split augmented Lagrangian method. We defined a regularization function based on the isotropic Ising model for the image support in the transform domain. In this way, we constrained that recovered image obeys desired characteristics imposed by the Ising model. A hard-thresholding rule is derived according to the adopted regularization function and the estimated image support mask in the transform domain. Since regularization, i.e., application of the hard-thresholding rule on image coefficients, is performed on the lattice structure defined by the Ising model, we named the algorithm lattice split augmented Lagrangian (LaSAL).

An extension of LaSAL with compound regularization which introduces the TV norm besides the MRF-based regularization function is considered. This results in an extended version of the LaSAL method named LaSAL2. The developed algorithms improve upon the constrained split augmented Lagrangian shrinkage algorithm (C-SALSA) in MRI reconstruction and they also outperform the earlier lattice split Bregman method (LaSB) which also utilized MRF signal prior. The proposed methods achieved much better reconstruction performances in comparison with the state-of-the-art approach that involved a wavelet-tree based model for the relations among image coefficients.

In the same chapter, we further considered a more general MRF model for the signal support and a fast composite splitting algorithm (FCSA) as an optimization framework in which we incorporated the proposed MRF-based regularization. In relation to LaSAL and LaSAL2 methods we now used anisotropic instead of the isotropic Ising model for modeling the support of image coefficients and we allow different model parameters for each image subband in the transformation domain. Introducing a more general MRF model is supported by the efficient estimation procedure of its parameters for each subband of coefficients. As well as in LaSAL2 where we considered MRF-based regularization in combination with TV. With the estimated support of image coefficients for each subband, we derived a new soft-thresholding rule instead of the hard-thresholding rule used in LaSAL and LaSAL2. Since the chosen optimization framework FCSA may go with and without acceleration steps, we proposed two methods composite splitting lattice with TV regularization (CSLaTV) and its accelerated version FCSLaTV. The proposed methods with automated MRF parameter estimation show great performances comparable to the LaSAL2 and give the possibility to reconstruct images with various structures (e.g. which came from imaging different human anatomy) for which fixed MRF parameters during reconstruction isn't optimal.

In Chapter 5 we considered a greedy iterative procedure for MRI reconstruction where we incorporate the proposed MRF-based regularization. We developed a greedy algorithm with lattice regularization (GreeLa) which performs hard-thresholding rule on image coefficients based on their estimated support. GreeLa is an extension of the lattice matching pursuit (LaMP) method developed primarily for images that are sparse in the canonical domain. The proposed GreeLa can reconstruct both the compressible MR images as well as on originally sparse images. The simplified iterative procedure of the GreeLa method

has a smaller number of parameters that need to be initialized compared to optimization-based methods proposed in Chapter 4. GreeLa showed stable behavior through iterations in comparison with the related MRF-based method LaSB and for low sampling rates outperforms best-performing methods proposed in Chapter 4. However, algorithm simplicity results in somewhat poorer performances with the increase of sampling rates compared to optimization-based methods. For sampling rates below 30%, GreeLa achieves comparable reconstruction results with optimization-based methods. Stable image recovery is obtained after 25 up to 30 algorithm iterations which is the result of the greedy iterative procedure in combination with MRF-based signal prior. Although this fast algorithm convergence to the solution is desired characteristic, it results in limiting algorithm performances for higher sampling rates in comparison with optimization-based methods.

In order to get closer to the real application, we also extend our MRF-based reconstruction framework to parallel MR data (pMRI-CS) in Chapter 6. Multi-coil reconstruction was conducted through the joint framework which takes into account undersampled measurements from all of the available coils and their estimated sensitivity profile maps. The proposed algorithm which is derived using the same procedure as for LaSAL2, with the addition of necessary transformations due to different algorithm inputs (multi coil measurements and sensitivity maps), shows great performances compared to the current state-of-the-art methods for pMRI-CS.

The exposed analysis and results in this thesis prove that the crucial benefit in MR reconstruction came from the involved MRF prior in both greedy and optimization-based frameworks and accordingly derived regularization rules. The results also demonstrate the superior performance of the proposed algorithm in comparison to state-of-the-art methods, both in terms of quantitative performance measures and visually.

There is much room to optimize the computations in our proposed methods, especially regarding the inference procedure in the MRF-based support estimation. Belief propagation algorithms may be considered as well as various parallelization procedures to optimize the code. Involving parallelization opens new possibilities for updating order for sites or a block of sites during inference procedure without increasing computation time. As well, estimation of MRF parameters can be elevated on the level of sites and cliques in their neighborhood, thereby allowing using more general non-homogeneous MRF models for a signal

support.

The applicability of the developed MRF-based regularization in different imaging problems should also be analyzed. The iterative reconstruction procedure used for computer tomography imaging has a great resemblance with the MRI reconstruction procedure and can be one of the possible applications of the proposed methods in this thesis. This might help in the reduction of time in which the patient is exposed to X-rays during the CT imaging. A joint image demosaicking and denoising for multispectral imaging can also be an interesting application. Here MRF signal prior can be used during the interpolation procedure to recover the image spectral bands to the original resolution of the mosaic-snapshot multispectral image.

Another very promising way of regularizing image estimate during reconstruction procedure is the usage of deep learning prior i.e. pre-trained deep neural networks (DNN) for denoising purposes. This is a very popular research direction with promising results, and it opens possibilities for a new interpretation of our developed MRF-based regularizer with the construction of DNN architecture which will be equivalent to proposed MRF-based regularization. This will lead to new ways of improving regularization through the DNN framework. These aspects will be part of our future research.

# Bibliography


[Adcock 13] Ben Adcock, Anders C. Hansen, Clarice Poon & Bogdan Roman. *Breaking the coherence barrier: asymptotic incoherence and asymptotic sparsity in compressed sensing.* CoRR, vol. abs/1302.0561, 2013.

[Adcock 14] Ben Adcock, Anders C Hansen & Bogdan Roman. *The quest for optimal sampling: Computationally efficient, structure-exploiting measurements for compressed sensing.* arXiv preprint arXiv:1403.6540, 2014.

[Aelterman 10] Jan Aelterman, Hiep Quang Luong, Bart Goossens, Aleksandra Pižurica & Wilfried Philips. *COMPASS: a joint framework for parallel imaging and compressive sensing in MRI.* In Image Processing (ICIP), 2010 17th IEEE International Conference on, pages 1653–1656. IEEE, 2010.

[Aelterman 11] Jan Aelterman, Hiêp Quang Luong, Bart Goossens, Aleksandra Pižurica & Wilfried Philips. *Augmented Lagrangian based reconstruction of non-uniformly sub-Nyquist sampled MRI data.* Signal Processing, vol. 91, no. 12, pages 2731–2742, 2011.

[Aelterman 14a] Jan Aelterman. *Multiresolution models in image restoration and reconstruction with medical and other applications.* PhD thesis, Ghent University, 2014.

[Aelterman 14b] Jan Aelterman, Maarten Naeyaert, Shandra Gutierrez, Hiep Luong, Bart Goossens, Aleksandra Pižurica & Wilfried Philips. *Automatic high-bandwidth calibration and reconstruction of arbitrar-*

*ily sampled parallel MRI.* PloS one, vol. 9, no. 6, page e98937, 2014.

[Afonso 09] Manya V Afonso, José M Bioucas-Dias & Mario AT Figueiredo. *Image restoration with compound regularization using a Bregman iterative algorithm.* In SPARS'09-Signal Processing with Adaptive Sparse Structured Representations, 2009.

[Afonso 10] Manya V Afonso, José M Bioucas-Dias & Mário AT Figueiredo. *Fast image recovery using variable splitting and constrained optimization.* IEEE Trans. Image Process., vol. 19, no. 9, pages 2345–2356, 2010.

[Afonso 11] Manya V Afonso, José M Bioucas-Dias & Mário AT Figueiredo. *An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems.* IEEE Trans. Image Process., vol. 20, no. 3, pages 681–695, 2011.

[Amrein 77] WO Amrein & Anne-Marie Berthier. *On support properties of Lp-functions and their Fourier transforms.* Journal of Functional Analysis, vol. 24, no. 3, pages 258–267, 1977.

[Antonini 92] Marc Antonini, Michel Barlaud, Pierre Mathieu & Ingrid Daubechies. *Image coding using wavelet transform.* IEEE Transactions on image processing, vol. 1, no. 2, pages 205–220, 1992.

[Baraniuk 07] Richard G Baraniuk. *Compressive sensing [lecture notes].* IEEE signal processing magazine, vol. 24, no. 4, pages 118–121, 2007.

[Baraniuk 10] Richard G Baraniuk, Volkan Cevher & Michael B Wakin. *Low-dimensional models for dimensionality reduction and signal recovery: A geometric perspective.* Proceedings of the IEEE, vol. 98, no. 6, pages 959–971, 2010.

[Bastounis 15] Alexander Bastounis & Anders C Hansen. *On random and deterministic compressed sensing and the Restricted Isometry Property in levels.* In Sampling

Theory and Applications (SampTA), 2015 International Conference on, pages 297–301. IEEE, 2015.

[Beck 09] Amir Beck & Marc Teboulle. *A fast iterative shrinkage-thresholding algorithm for linear inverse problems.* SIAM Journal on Imaging Sciences, vol. 2, no. 1, pages 183–202, 2009.

[Besag 74] Julian Besag. *Spatial interaction and the statistical analysis of lattice systems.* Journal of the Royal Statistical Society. Series B (Methodological), pages 192–236, 1974.

[Besag 86] Julian Besag. *On the statistical analysis of dirty pictures.* Journal of the Royal Statistical Society. Series B (Methodological), pages 259–302, 1986.

[Bioucas-Dias 07] José M Bioucas-Dias & Mário AT Figueiredo. *A new TwIST: two-step iterative shrinkage/thresholding algorithms for image restoration.* IEEE Trans. Image Process., vol. 16, no. 12, pages 2992–3004, 2007.

[Bioucas-Dias 08] José M Bioucas-Dias & Mário AT Figueiredo. *An iterative algorithm for linear inverse problems with compound regularizers.* In 2008 15th IEEE International Conference on Image Processing, pages 685–688. IEEE, 2008.

[Blaimer 04] Martin Blaimer, Felix Breuer, Matthias Mueller, Robin M Heidemann, Mark A Griswold & Peter M Jakob. *SMASH, SENSE, PILS, GRAPPA: how to choose the optimal method.* Topics in Magnetic Resonance Imaging, vol. 15, no. 4, pages 223–236, 2004.

[Blaimer 16] Martin Blaimer, Marius Heim, Daniel Neumann, Peter M Jakob, Stephan Kannengiesser & Felix A Breuer. *Comparison of phase-constrained parallel MRI approaches: Analogies and differences.* Magnetic resonance in medicine, vol. 75, no. 3, pages 1086–1099, 2016.

[Bloch 46] Felix Bloch. *Nuclear induction.* Physical review, vol. 70, no. 7-8, page 460, 1946.

[Blumensath 09] Thomas Blumensath & Mike E Davies. *Iterative hard thresholding for compressed sensing.* Applied and Computational Harmonic Analysis, vol. 27, no. 3, pages 265–274, 2009.

[Blumensath 10] Thomas Blumensath & Mike E Davies. *Normalized iterative hard thresholding: Guaranteed stability and performance.* Selected Topics in Signal Processing, IEEE Journal of, vol. 4, no. 2, pages 298–309, 2010.

[Blumensath 12] Thomas Blumensath. *Accelerated iterative hard thresholding.* Signal Processing, vol. 92, no. 3, pages 752–756, 2012.

[Bregman 67] Lev M Bregman. *The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming.* USSR computational mathematics and mathematical physics, vol. 7, no. 3, pages 200–217, 1967.

[Brown 14] Robert W Brown, E Mark Haacke, Y-C Norman Cheng, Michael R Thompson & Ramesh Venkatesan. Magnetic resonance imaging: physical principles and sequence design. John Wiley & Sons, 2014.

[Bruce 02] Lori Mann Bruce, Cliff H Koger & Jiang Li. *Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction.* IEEE Transactions on geoscience and remote sensing, vol. 40, no. 10, pages 2331–2338, 2002.

[Bruckstein 09] Alfred M Bruckstein, David L Donoho & Michael Elad. *From sparse solutions of systems of equations to sparse modeling of signals and images.* SIAM review, vol. 51, no. 1, pages 34–81, 2009.

[Bydder 05] Mark Bydder & Matthew D Robson. *Partial Fourier partially parallel imaging.* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, vol. 53, no. 6, pages 1393–1401, 2005.

[Candès 99] Emmanuel J Candès & David L Donoho. *Ridgelets: A key to higher-dimensional intermittency?* Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, vol. 357, no. 1760, pages 2495–2509, 1999.

[Candès 04] Emmanuel J Candès & David L Donoho. *New tight frames of curvelets and optimal representations of objects with piecewise C2 singularities.* Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences, vol. 57, no. 2, pages 219–266, 2004.

[Candès 06a] Emmanuel J Candès*et al. Compressive sampling.* In Proceedings of the international congress of mathematicians, volume 3, pages 1433–1452. Madrid, Spain, 2006.

[Candès 06b] Emmanuel J Candès*et al. Compressive sampling.* In Proceedings of the international congress of mathematicians, volume 3, pages 1433–1452. Madrid, Spain, 2006.

[Candes 06c] Emmanuel J Candes & Justin Romberg. *Quantitative robust uncertainty principles and optimally sparse decompositions.* Foundations of Computational Mathematics, vol. 6, no. 2, pages 227–254, 2006.

[Candès 06d] Emmanuel J Candès, Justin Romberg & Terence Tao. *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information.* IEEE Transactions on information theory, vol. 52, no. 2, pages 489–509, 2006.

[Candes 06e] Emmanuel J Candes, Justin K Romberg & Terence Tao. *Stable signal recovery from incomplete and inaccurate measurements.* Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences, vol. 59, no. 8, pages 1207–1223, 2006.

[Candes 06f] Emmanuel J Candes & Terence Tao. *Near-optimal signal recovery from random projections: Universal encoding strategies?* IEEE transactions on information theory, vol. 52, no. 12, pages 5406–5425, 2006.

[Candes 08] Emmanuel J Candes. *The restricted isometry property and its implications for compressed sensing.* Comptes rendus mathematique, vol. 346, no. 9-10, pages 589–592, 2008.

[Candes 11] Emmanuel J Candes, Yonina C Eldar, Deanna Needell & Paige Randall. *Compressed sensing with coherent and redundant dictionaries.* Applied and Computational Harmonic Analysis, vol. 31, no. 1, pages 59–73, 2011.

[Cevher 09a] Volkan Cevher. *Learning with compressible priors.* In Advances in Neural Information Processing Systems, pages 261–269, 2009.

[Cevher 09b] Volkan Cevher, Marco F Duarte, Chinmay Hegde & Richard Baraniuk. *Sparse signal recovery using Markov random fields.* In Advances in Neural Information Processing Systems, pages 257–264, 2009.

[Cevher 10a] Volkan Cevher, Piotr Indyk, Lawrence Carin & Richard G Baraniuk. *Sparse signal recovery and acquisition with graphical models.* IEEE Signal Processing Magazine, vol. 27, no. 6, pages 92–103, 2010.

[Cevher 10b] Volkan Cevher, Piotr Indyk, Lawrence Carin & Richard G Baraniuk. *Sparse signal recovery and acquisition with graphical models.* IEEE Signal Process. Mag., vol. 27, no. 6, pages 92–103, 2010.

[Chambolle 04] Antonin Chambolle. *An algorithm for total variation minimization and applications.* Journal of Mathematical imaging and vision, vol. 20, no. 1-2, pages 89–97, 2004.

[Chen 98] Scott Shaobing Chen, David L Donoho & Michael A Saunders. *Atomic decomposition by basis pursuit.* SIAM Journal on Scientific Computing, vol. 20, no. 1, pages 33–61, 1998.

[Chen 12] Chen Chen & Junzhou Huang. *Compressive sensing MRI with wavelet tree sparsity.* In Advances in neural information processing systems, pages 1115–1123, 2012.

[Chen 14] Chen Chen & Junzhou Huang. *Exploiting the wavelet structure in compressed sensing MRI.* Magnetic Resonance Imaging, vol. 32, no. 10, pages 1377–1389, 2014.

[Cheng 16] Hong Cheng. Sparse representation, modeling and learning in visual recognition. Springer, 2016.

[Chui 16] Charles K Chui. An introduction to wavelets. Elsevier, 2016.

[Chun 15] Il Yong Chun, Ben Adcock & Thomas M Talavage. *Efficient compressed sensing SENSE pMRI reconstruction with joint sparsity promotion.* IEEE transactions on medical imaging, vol. 35, no. 1, pages 354–368, 2015.

[Cline 01] Harvey Ellis Cline & Thomas Richard Anthony. *Magnetic resonance imaging with interleaved fibonacci spiral scanning*, August 28 2001. US Patent 6,281,681.

[Cohen 96] Albert Cohen & Jelena Kovacevic. *Wavelets: The mathematical background.* Proceedings of the IEEE, vol. 84, no. 4, pages 514–522, 1996.

[Cohen 09] Albert Cohen, Wolfgang Dahmen & Ronald DeVore. *Compressed sensing and best k-term approximation.* Journal of the American mathematical society, vol. 22, no. 1, pages 211–231, 2009.

[Coifman 95] Ronald R Coifman & David L Donoho. *Translation-invariant de-noising.* In Wavelets and statistics, pages 125–150. Springer, 1995.

[Combettes 05] Patrick L Combettes & Valérie R Wajs. *Signal recovery by proximal forward-backward splitting.* Multiscale Modeling & Simulation, vol. 4, no. 4, pages 1168–1200, 2005.

[Courant 94] Richard Courant*et al.* *Variational methods for the solution of problems of equilibrium and vibrations.* Lecture Notes in Pure and Applied Mathematics, pages 1–1, 1994.

[Dabov 07] Kostadin Dabov, Alessandro Foi & Karen Egiazarian. *Video denoising by sparse 3D transform-domain collaborative filtering.* In 2007 15th European Signal Processing Conference, pages 145–149. IEEE, 2007.

[Dai 09] Wei Dai & Olgica Milenkovic. *Subspace pursuit for compressive sensing signal reconstruction.* IEEE transactions on Information Theory, vol. 55, no. 5, pages 2230–2249, 2009.

[Danielyan 11] Aram Danielyan, Vladimir Katkovnik & Karen Egiazarian. *BM3D frames and variational image deblurring.* IEEE Transactions on Image Processing, vol. 21, no. 4, pages 1715–1728, 2011.

[Daubechies 88] Ingrid Daubechies. *Orthonormal bases of compactly supported wavelets.* Communications on pure and applied mathematics, vol. 41, no. 7, pages 909–996, 1988.

[Daubechies 92] Ingrid Daubechies. Ten lectures on wavelets, volume 61. Siam, 1992.

[Daubechies 04] Ingrid Daubechies, Michel Defrise & Christine De Mol. *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint.* Communications on pure and applied mathematics, vol. 57, no. 11, pages 1413–1457, 2004.

[Deka 18] Bhabesh Deka, Sumit Datta & Sanjeev Handique. *Wavelet tree support detection for compressed sensing MRI reconstruction.* IEEE Signal Processing Letters, vol. 25, no. 5, pages 730–734, 2018.

[Den Dekker 14] AJ Den Dekker & J Sijbers. *Data distributions in magnetic resonance images: a review.* Physica Medica, vol. 30, no. 7, pages 725–741, 2014.

[Deng 11] Chun Yuan Deng. *A generalization of the Sherman–Morrison–Woodbury formula.* Applied Mathematics Letters, vol. 24, no. 9, pages 1561–1564, 2011.

[Do 05] Minh N Do & Martin Vetterli. *The contourlet transform: an efficient directional multiresolution image representation.* IEEE Transactions on image processing, vol. 14, no. 12, pages 2091–2106, 2005.

[Donoho 95] David L Donoho. *De-noising by soft-thresholding.* IEEE transactions on information theory, vol. 41, no. 3, pages 613–627, 1995.

[Donoho 06a] David L Donoho. *Compressed sensing.* IEEE Transactions on information theory, vol. 52, no. 4, pages 1289–1306, 2006.

[Donoho 06b] David L Donoho. *Compressed sensing.* IEEE Trans. Inf. Theory, vol. 52, no. 4, pages 1289–1306, 2006.

[Donoho 06c] David Leigh Donoho, Iddo Drori, Yaakov Tsaig & Jean-Luc Starck. Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit. Department of Statistics, Stanford University, 2006.

[Douglas 56] Jim Douglas & Henry H Rachford. *On the numerical solution of heat conduction problems in two and three space variables.* Transactions of the American mathematical Society, vol. 82, no. 2, pages 421–439, 1956.

[Easley 08] Glenn Easley, Demetrio Labate & Wang-Q Lim. *Sparse directional image representations using the discrete shearlet transform.* Applied and Computational Harmonic Analysis, vol. 25, no. 1, pages 25–46, 2008.

[Easley 09] Glenn R Easley, Demetrio Labate & Flavia Colonna. *Shearlet-based total variation diffusion for denoising.* IEEE Transactions on Image processing, vol. 18, no. 2, pages 260–268, 2009.

[Eckstein 92] Jonathan Eckstein & Dimitri P Bertsekas. *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators.* Mathematical Programming, vol. 55, no. 1-3, pages 293–318, 1992.

[Eksioglu 18] Ender M Eksioglu & A Korhan Tanc. *Denoising AMP for MRI reconstruction: BM3D-AMP-MRI.* SIAM Journal on Imaging Sciences, vol. 11, no. 3, pages 2090–2109, 2018.

[Elad 07] Michael Elad, Boaz Matalon & Michael Zibulevsky. *Coordinate and subspace optimization methods for linear least squares with non-quadratic regularization.* Applied and Computational Harmonic Analysis, vol. 23, no. 3, pages 346–367, 2007.

[Esser 09] Ernie Esser. *Applications of Lagrangian-based alternating direction methods and connections to split Bregman.* CAM report, vol. 9, page 31, 2009.

[Fessler 03] Jeffrey A Fessler & Bradley P Sutton. *Nonuniform fast Fourier transforms using min-max interpolation.* 2003.

[Fessler 10] Jeffrey A Fessler. *Model-based image reconstruction for MRI.* IEEE Signal Processing Magazine, vol. 27, no. 4, pages 81–89, 2010.

[Figueiredo 07] Mário AT Figueiredo, Robert D Nowak & Stephen J Wright. *Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems.* IEEE Journal of selected topics in signal processing, vol. 1, no. 4, pages 586–597, 2007.

[Fowler 05] James E Fowler. *The redundant discrete wavelet transform and additive noise.* IEEE Signal Processing Letters, vol. 12, no. 9, pages 629–632, 2005.

[Gabay 75] Daniel Gabay & Bertrand Mercier. A dual algorithm for the solution of non linear variational problems via finite element approximation. Institut de recherche d'informatique et d'automatique, 1975.

[Gabor 46] Dennis Gabor. *Theory of communication. Part 1: The analysis of information.* Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering, vol. 93, no. 26, pages 429–441, 1946.

[Gaetan 10] Carlo Gaetan & Xavier Guyon. Spatial statistics and modeling, volume 81. Springer, 2010.

[Ghofrani 15] Sedigheh Ghofrani. *Comparing Nonsubsampled Wavelet, Contourlet and Shearlet Transforms for Ultrasound Image Despeckling.* International Journal of Image, Graphics and Signal Processing, vol. 7, no. 2, page 15, 2015.

[Gibert 14] Xavier Gibert, Vishal M Patel, Demetrio Labate & Rama Chellappa. *Discrete shearlet transform on GPU with applications in anomaly detection and denoising.* EURASIP Journal on Advances in Signal Processing, vol. 2014, no. 1, page 64, 2014.

[Glowinski 75] Roland Glowinski & A Marroco. *Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires.* Revue française d'automatique, informatique, recherche opérationnelle. Analyse numérique, vol. 9, no. R2, pages 41–76, 1975.

[Goldstein 09] Tom Goldstein & Stanley Osher. *The split Bregman method for $\ell_1$-regularized problems.* SIAM Journal on Imaging Sciences, vol. 2, no. 2, pages 323–343, 2009.

[Goossens 09] Bart Goossens, Jan Aelterman, Hiêp Luong, Aleksandra Pižurica & Wilfried Philips. *Efficient design of a low redundant discrete shearlet transform.* In IEEE Internat. Workshop on Local and Non-Local Approximation in Image Processing, LNLA, pages 112–124, 2009.

[Greiser 03] Andreas Greiser & Markus von Kienlin. *Efficient k-space sampling by density-weighted phase-encoding.*

Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, vol. 50, no. 6, pages 1266–1275, 2003.

[Griswold 02] Mark A Griswold, Peter M Jakob, Robin M Heidemann, Mathias Nittka, Vladimir Jellus, Jianmin Wang, Berthold Kiefer & Axel Haase. *Generalized autocalibrating partially parallel acquisitions (GRAPPA).* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, vol. 47, no. 6, pages 1202–1210, 2002.

[Grossmann 84] Alexander Grossmann & Jean Morlet. *Decomposition of Hardy functions into square integrable wavelets of constant shape.* SIAM journal on mathematical analysis, vol. 15, no. 4, pages 723–736, 1984.

[Guo 95] H Guo, M Lang, JE Odegard & CS Burrus. *Nonlinear processing of a shift-invariant DWT for noise reduction and compression.* In Proceedings of the International Conference on Digital Signal Processing, pages 332–337, 1995.

[Guo 06] Kanghui Guo, Gitta Kutyniok & Demetrio Labate. *Sparse multidimensional representations using anisotropic dilation and shear operators*, 2006.

[Guo 07] Kanghui Guo & Demetrio Labate. *Optimally sparse multidimensional representation using shearlets.* SIAM journal on mathematical analysis, vol. 39, no. 1, pages 298–318, 2007.

[Haar 10] Alfred Haar. *Zur theorie der orthogonalen funktionensysteme.* Mathematische Annalen, vol. 69, no. 3, pages 331–371, 1910.

[Hahn 50] Erwin L Hahn. *Spin echoes.* Physical review, vol. 80, no. 4, page 580, 1950.

[Haldar 14] Justin P Haldar. *Low-Rank Modeling of Local k-Space Neighborhoods (LORAKS) for Constrained*

*MRI.* IEEE transactions on medical imaging, vol. 33, no. 3, pages 668–681, 2014.

[Haldar 16] Justin P Haldar & Jingwei Zhuo. *P-LORAKS: Low-rank modeling of local k-space neighborhoods with parallel imaging data.* Magnetic resonance in medicine, vol. 75, no. 4, pages 1499–1514, 2016.

[He 09] Lihan He & Lawrence Carin. *Exploiting Structure in Wavelet-Based Bayesian Compressive Sensing.* IEEE Trans. Signal Process., vol. 57, no. 9, pages 3488–3497, 2009.

[He 10] Lihan He, Haojun Chen & Lawrence Carin. *Tree-structured compressive sensing with variational Bayesian analysis.* IEEE Signal Processing Letters, vol. 17, no. 3, pages 233–236, 2010.

[Hestenes 69] Magnus R Hestenes. *Multiplier and gradient methods.* Journal of optimization theory and applications, vol. 4, no. 5, pages 303–320, 1969.

[Hinman 84] B Hinman, Jared Bernstein & D Staelin. *Short-space Fourier transform image processing.* In ICASSP'84. IEEE International Conference on Acoustics, Speech, and Signal Processing, volume 9, pages 166–169. IEEE, 1984.

[Holschneider 90] Matthias Holschneider, Richard Kronland-Martinet, Jean Morlet & Ph Tchamitchian. *A real-time algorithm for signal analysis with the help of the wavelet transform.* In Wavelets, pages 286–297. Springer, 1990.

[Hornak 06] Joseph P Hornak. *The basics of MRI.* http://www.cis. rit. edu/htbooks/mri, 2006.

[Huang 11a] Junzhou Huang, Shaoting Zhang, Hongsheng Li & Dimitris Metaxas. *Composite splitting algorithms for convex optimization.* Computer Vision and Image Understanding, vol. 115, no. 12, pages 1610–1622, 2011.

[Huang 11b] Junzhou Huang, Shaoting Zhang & Dimitris Metaxas. *Efficient MR image reconstruction for compressed MR imaging.* Medical Image Analysis, vol. 15, no. 5, pages 670–679, 2011.

[Huang 12] Junzhou Huang & Fei Yang. *Compressed magnetic resonance imaging based on wavelet sparsity and nonlocal total variation.* In 9th IEEE Internat. Symp. on Biomedical Imaging (ISBI), pages 968–971, 2012.

[Huang 14] Yue Huang, John Paisley, Qin Lin, Xinghao Ding, Xueyang Fu & Xiao-Ping Zhang. *Bayesian nonparametric dictionary learning for compressed sensing MRI.* IEEE Trans. Image Process., vol. 23, no. 12, pages 5007–5019, 2014.

[Jacques 11] Laurent Jacques, Laurent Duval, Caroline Chaux & Gabriel Peyré. *A panorama on multiscale geometric representations, intertwining spatial, directional and frequency selectivity.* Signal Processing, vol. 91, no. 12, pages 2699–2730, 2011.

[Jin 16] Kyong Hwan Jin, Dongwook Lee & Jong Chul Ye. *A general framework for compressed sensing and parallel MRI using annihilating filter based low-rank Hankel matrix.* IEEE Transactions on Computational Imaging, vol. 2, no. 4, pages 480–495, 2016.

[Johnson 09] Kenneth O Johnson & James G Pipe. *Convolution kernel design and efficient algorithm for sampling density correction.* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, vol. 61, no. 2, pages 439–447, 2009.

[Kim 17] Tae Hyung Kim, Kawin Setsompop & Justin P Haldar. *LORAKS makes better SENSE: Phase-constrained partial fourier SENSE reconstruction without phase calibration.* Magnetic resonance in medicine, vol. 77, no. 3, pages 1021–1035, 2017.

[Kingsbury 01] Nick Kingsbury. *Complex wavelets for shift invariant analysis and filtering of signals.* Applied and computational harmonic analysis, vol. 10, no. 3, pages 234–253, 2001.

[Kirkpatrick 84] Scott Kirkpatrick. *Optimization by simulated annealing: Quantitative studies.* J. Stat. Phys., vol. 34, no. 5-6, pages 975–986, 1984.

[Kolmogorov 04] Vladimir Kolmogorov & Ramin Zabih. *What energy functions can be minimized via graph cuts ?* IEEE Trans. Pattern Anal. Mach. Intell, vol. 26, pages 65–81, 2004.

[Korosec 96] Frank R Korosec, Richard Frayne, Thomas M Grist & Charles A Mistretta. *Time-resolved contrast-enhanced 3D MR angiography.* Magnetic Resonance in Medicine, vol. 36, no. 3, pages 345–351, 1996.

[Kotelnikov 33] Vladimir Aleksandrovich Kotelnikov. *On the transmission capacity of the ether and of cables in electrical communications.* In Proceedings of the first All-Union Conference on the technological reconstruction of the communications sector and the development of low-current engineering. Moscow. Citeseer, 1933.

[Kuperman 00] Vadim Kuperman. Magnetic resonance imaging: physical principles and applications. Elsevier, 2000.

[Kyrillidis 15] Anastasios Kyrillidis, Luca Baldassarre, Marwa El Halabi, Quoc Tran-Dinh & Volkan Cevher. *Structured Sparsity: Discrete and Convex Approaches.* In Compressed Sensing and its Applications, pages 341–387. Springer, 2015.

[Labate 05] Demetrio Labate, Wang-Q Lim, Gitta Kutyniok & Guido Weiss. *Sparse multidimensional representation using shearlets.* In Wavelets XI, volume 5914, page 59140U. International Society for Optics and Photonics, 2005.

[Lai 10] P Lai, M Lustig, AC Brau, S Vasanawala, PJ Beatty & M Alley. *Efficient L1SPIRiT reconstruction (ESPIRiT) for highly accelerated 3D volumetric MRI with parallel imaging and compressed sensing.* In Proceedings of the Joint Annual Meeting ISMRM-ESMRMB, Stockholm, Sweden, volume 345, 2010.

[Lai 16] Zongying Lai, Xiaobo Qu, Yunsong Liu, Di Guo, Jing Ye, Zhifang Zhan & Zhong Chen. *Image reconstruction of compressed sensing MRI using graph-based redundant wavelet transform.* Medical image analysis, vol. 27, pages 93–104, 2016.

[Lang 96] Markus Lang, Haitao Guo, Jan E Odegard, C Sidney Burrus & Raymond O Wells. *Noise reduction using an undecimated discrete wavelet transform.* IEEE Signal Processing Letters, vol. 3, no. 1, pages 10–12, 1996.

[Lauterbur 73] Paul C Lauterbur*et al.* *Image formation by induced local interactions: examples employing nuclear magnetic resonance.* 1973.

[Le Pennec 05] Erwan Le Pennec & Stéphane Mallat. *Sparse geometric image representations with bandelets.* IEEE transactions on image processing, vol. 14, no. 4, pages 423–438, 2005.

[Li 09] Stan Z Li. Markov random field modeling in image analysis. Springer Science & Business Media, 2009.

[Liang 00] Zhi-Pei Liang & Paul C Lauterbur. Principles of magnetic resonance imaging: a signal processing perspective. SPIE Optical Engineering Press, 2000.

[Liang 09] Dong Liang, Bo Liu, Jiunjie Wang & Leslie Ying. *Accelerating SENSE using compressed sensing.* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, vol. 62, no. 6, pages 1574–1584, 2009.

[Lin 00] Jing Lin & Liangsheng Qu. *Feature extraction based on Morlet wavelet and its application for mechanical fault diagnosis.* Journal of sound and vibration, vol. 234, no. 1, pages 135–148, 2000.

[Lions 79] Pierre-Louis Lions & Bertrand Mercier. *Splitting algorithms for the sum of two nonlinear operators.* SIAM Journal on Numerical Analysis, vol. 16, no. 6, pages 964–979, 1979.

[Liu 08] Bo Liu, Yi Ming Zou & Leslie Ying. *SparseSENSE: application of compressed sensing in parallel MRI.* In Information Technology and Applications in Biomedicine, 2008. ITAB 2008. International Conference on, pages 127–130. IEEE, 2008.

[Liu 16] Yunsong Liu, Zhifang Zhan, Jianfeng Cai, Di Guo, Zhong Chen & Xiaobo Qu. *Projected iterative soft-thresholding algorithm for tight frames in compressed sensing magnetic resonance imaging.* IEEE Trans. Med. Imag, vol. 35, no. 9, pages 2130–2140, 2016.

[Loh 17] Po-Ling Loh*et al.* *Statistical consistency and asymptotic normality for high-dimensional robust M-estimators.* The Annals of Statistics, vol. 45, no. 2, pages 866–896, 2017.

[Lustig 07] Michael Lustig, David Donoho & John M Pauly. *Sparse MRI: The application of compressed sensing for rapid MR imaging.* Magnetic Resonance in Medicine, vol. 58, no. 6, pages 1182–1195, 2007.

[Lustig 08] Michael Lustig, David L Donoho, Juan M Santos & John M Pauly. *Compressed sensing MRI.* IEEE Signal Process. Mag., vol. 25, no. 2, pages 72–82, 2008.

[Lustig 10] Michael Lustig & John M Pauly. *SPIRiT: iterative self-consistent parallel imaging reconstruction from arbitrary k-space.* Magnetic resonance in medicine, vol. 64, no. 2, pages 457–471, 2010.

[Ma 08] Shiqian Ma, Wotao Yin, Yin Zhang & Amit Chakraborty. *An efficient algorithm for compressed MR imaging using total variation and wavelets.* In IEEE Conf. on Computer Vision and Pattern Recognition, CVPR, pages 1–8, 2008.

[Ma 17] Junjie Ma & Li Ping. *Orthogonal amp.* IEEE Access, vol. 5, pages 2020–2033, 2017.

[Maleki 10] Mohammad Ali Maleki. Approximate message passing algorithms for compressed sensing. Stanford University, 2010.

[Mallat 89a] Stephane G Mallat. *Multiresolution approximations and wavelet orthonormal bases of $L^2$ (R).* Transactions of the American mathematical society, vol. 315, no. 1, pages 69–87, 1989.

[Mallat 89b] Stephane G Mallat. *A theory for multiresolution signal decomposition: the wavelet representation.* IEEE transactions on pattern analysis and machine intelligence, vol. 11, no. 7, pages 674–693, 1989.

[Mallat 91] Stephane Mallat. *Zero-crossings of a wavelet transform.* IEEE Transactions on Information theory, vol. 37, no. 4, pages 1019–1033, 1991.

[Mallat 92] Stephane Mallat & Wen Liang Hwang. *Singularity detection and processing with wavelets.* IEEE transactions on information theory, vol. 38, no. 2, pages 617–643, 1992.

[Mallat 96] Stephane Mallat. *Wavelets for a vision.* Proceedings of the IEEE, vol. 84, no. 4, pages 604–614, 1996.

[Mallat 99] Stéphane Mallat. A wavelet tour of signal processing. Elsevier, 1999.

[Mansfield 77] Peter Mansfield. *Multi-planar image formation using NMR spin echoes.* Journal of Physics C: Solid State Physics, vol. 10, no. 3, page L55, 1977.

[Marseille 96] GJ Marseille, R De Beer, M Fuderer, AF Mehlkopf & D Van Ormondt. *Nonuniform phase-encode distributions for MRI scan time reduction.* Journal of Magnetic Resonance, vol. 111, pages 70–75, 1996.

[McGibney 93] G McGibney, MR Smith, ST Nichols & A Crawley. *Quantitative evaluation of several partial Fourier reconstruction algorithms used in MRI.* Magnetic resonance in medicine, vol. 30, no. 1, pages 51–59, 1993.

[Metzler 16] Christopher A Metzler, Arian Maleki & Richard G Baraniuk. *From denoising to compressed sensing.* IEEE Transactions on Information Theory, vol. 62, no. 9, pages 5117–5144, 2016.

[Meyer 89] Yves Meyer. *Orthonormal wavelets.* In Wavelets, pages 21–37. Springer, 1989.

[Meyer 97] François G Meyer & Ronald R Coifman. *Brushlets: a tool for directional image analysis and image compression.* Applied and computational harmonic analysis, vol. 4, no. 2, pages 147–187, 1997.

[Mistretta 06] Charles A Mistretta, O Wieben, J Velikina, W Block, J Perry, Yijing Wu, K Johnson & Yan Wu. *Highly constrained backprojection for time-resolved MRI.* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, vol. 55, no. 1, pages 30–40, 2006.

[Morlet 82] Jean Morlet, G Arens, E Fourgeau & D Glard. *Wave propagation and sampling theory Part I: Complex signal and scattering in multilayered media.* Geophysics, vol. 47, no. 2, pages 203–221, 1982.

[Moussouris 74] John Moussouris. *Gibbs and Markov random systems with constraints.* Journal of statistical physics, vol. 10, no. 1, pages 11–33, 1974.

[Murphy 99] Kevin P Murphy, Yair Weiss & Michael I Jordan. *Loopy belief propagation for approximate inference:*

*An empirical study.* In Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence, pages 467–475. Morgan Kaufmann Publishers Inc., 1999.

[Nason 95] Guy P Nason & Bernard W Silverman. *The stationary wavelet transform and some statistical applications.* In Wavelets and statistics, pages 281–299. Springer, 1995.

[Needell 09a] Deanna Needell & Joel A Tropp. *CoSaMP: Iterative signal recovery from incomplete and inaccurate samples.* Applied and computational harmonic analysis, vol. 26, no. 3, pages 301–321, 2009.

[Needell 09b] Deanna Needell & Joel A Tropp. *CoSaMP: Iterative signal recovery from incomplete and inaccurate samples.* Applied and Computational Harmonic Analysis, vol. 26, no. 3, pages 301–321, 2009.

[Niesen 09] Urs Niesen, Devavrat Shah & Gregory W Wornell. *Adaptive alternating minimization algorithms.* IEEE Transactions on Information Theory, vol. 55, no. 3, pages 1423–1429, 2009.

[Nocedal 06] Jorge Nocedal & Stephen J Wright. *Numerical optimization 2nd*, 2006.

[Nowogrodzki 18] Anna Nowogrodzki. *The world's strongest MRI machines are pushing human imaging to new limits.*, 2018.

[Nyquist 28] Harry Nyquist. *Certain topics in telegraph transmission theory.* Transactions of the American Institute of Electrical Engineers, vol. 47, no. 2, pages 617–644, 1928.

[Osher 05] Stanley Osher, Martin Burger, Donald Goldfarb, Jinjun Xu & Wotao Yin. *An iterative regularization method for total variation-based image restoration.* Multiscale Modeling & Simulation, vol. 4, no. 2, pages 460–489, 2005.

[Panić 16a] M. Panić, J. Aelterman, V. Crnojević & A. Pižurica. *Compressed sensing in MRI with a Markov random field prior for spatial clustering of subband coefficients.* In 2016 24th European Signal Processing Conference (EUSIPCO), pages 562–566, Aug 2016.

[Panić 16b] Marko Panić, Dejan Vukobratović, Vladimir Crnojević & Aleksandra Pižurica. *Sparse MRI with a Markov Random Field Prior for the Subband Coefficients.* In Proceedings of the third "international Traveling Workshop on Interactions between Sparse models and Technology" (iTWIST'16), pages 56–58, 2016.

[Panić 17a] Marko Panić, Jan Aelterman, Vladimir Crnojević & Aleksandra Pižurica. *Sparse Recovery in Magnetic Resonance Imaging With a Markov Random Field Prior.* IEEE transactions on medical imaging, vol. 36, no. 10, pages 2104–2115, 2017.

[Panić 17b] Marko Panić, Dejan Vukobratović, Vladimir Crnojević & Aleksandra Pižurica. *Greedy MRI reconstruction using Markov Random Field prior.* In Proceedings of the IEICE Information and Communication Technology Forum (ICTF 2017). Polish Association of Telecommunication Engineer, 2017.

[Panić 19] Marko Panić, Jan Aelterman, Vladimir Crnojevic & Aleksandra Pizurica. *Multi-coil magnetic resonance imaging reconstruction with a Markov Random Field prior.* In SPIE Medical Imaging, 2019.

[Pejoski 15] Slavche Pejoski, Venceslav Kafedziski & Dušan Gleich. *Compressed sensing MRI using discrete nonseparable shearlet transform and FISTA.* IEEE Signal Process. Lett., vol. 22, no. 10, pages 1566–1570, 2015.

[Peters 06] Dana C Peters, Pratik Rohatgi, René M Botnar, Susan B Yeon, Kraig V Kissinger & Warren J Manning. *Characterizing radial undersampling artifacts for cardiac applications.* Magnetic Resonance

in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, vol. 55, no. 2, pages 396–403, 2006.

[Pittner 99] Stefan Pittner & Sagar V. Kamarthi. *Feature extraction from wavelet coefficients for pattern recognition tasks.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 21, no. 1, pages 83–88, 1999.

[Pižurica 02] Aleksandra Pižurica, Wilfried Philips, Ignace Lemahieu & Marc Acheroy. *A joint inter-and intrascale statistical model for Bayesian wavelet based image denoising.* IEEE Trans. Image Process., vol. 11, no. 5, pages 545–557, 2002.

[Pizurica 06] Aleksandra Pizurica & Wilfried Philips. *Estimating the probability of the presence of a signal of interest in multiresolution single-and multiband image denoising.* IEEE Transactions on image processing, vol. 15, no. 3, pages 654–665, 2006.

[Pižurica 11] Aleksandra Pižurica, Jan Aelterman, Funing Bai, Sam Vanloocke, Quang Luong, Bart Goossens & Wilfried Philips. *On structured sparsity and selected applications in tomographic imaging.* In SPIE Conference on Wavelets and Sparsity XIV, volume 8138, pages 81381D–1–12, 2011.

[Portilla 03] Javier Portilla, Vasily Strela, Martin J Wainwright & Eero P Simoncelli. *Image denoising using scale mixtures of Gaussians in the wavelet domain.* IEEE Trans Image Processing, vol. 12, no. 11, 2003.

[Protter 07] Matan Protter & Michael Elad. *Sparse and redundant representations and motion-estimation-free algorithm for video denoising.* In Wavelets XII, volume 6701, page 67011D. International Society for Optics and Photonics, 2007.

[Pruessmann 99] Klaas P Pruessmann, Markus Weiger, Markus B Scheidegger & Peter Boesiger. *SENSE: sensitiv-*

*ity encoding for fast MRI.* Magnetic resonance in medicine, vol. 42, no. 5, pages 952–962, 1999.

[Pruessmann 01] Klaas P Pruessmann, Markus Weiger, Peter Börnert & Peter Boesiger. *Advances in sensitivity encoding with arbitrary k-space trajectories.* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, vol. 46, no. 4, pages 638–651, 2001.

[Pruessmann 06] Klaas P Pruessmann. *Encoding and reconstruction in parallel MRI.* NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magnetic Resonance In vivo, vol. 19, no. 3, pages 288–299, 2006.

[Purcell 46] Edward M Purcell, H Co Torrey & Robert V Pound. *Resonance absorption by nuclear magnetic moments in a solid.* Physical review, vol. 69, no. 1-2, page 37, 1946.

[Qu 10] Xiaobo Qu, Weiru Zhang, Di Guo, Congbo Cai, Shuhui Cai & Zhong Chen. *Iterative thresholding compressed sensing MRI based on contourlet transform.* Inverse Probl. Sci. Eng., vol. 18, no. 6, pages 737–758, 2010.

[Qu 14] Xiaobo Qu, Yingkun Hou, Fan Lam, Di Guo, Jianhui Zhong & Zhong Chen. *Magnetic resonance image reconstruction from undersampled measurements using a patch-based nonlocal operator.* Medical image analysis, vol. 18, no. 6, pages 843–856, 2014.

[Rabi 38] Isidor Isaac Rabi, Jerrold R Zacharias, Sidney Millman & Polykarp Kusch. *A new method of measuring nuclear magnetic moment.* Physical Review, vol. 53, no. 4, page 318, 1938.

[Rajani 12] S. R. Rajani & M. Ramasubba Reddy. *An iterative hard thresholding algorithm for CS MRI.* In SPIE Medical Imaging, pages 83143W1–83143W7, 2012.

[Ramani 10] Sathish Ramani & Jeffrey A Fessler. *An accelerated iterative reweighted least squares algorithm for compressed sensing MRI.* In 2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pages 257–260. IEEE, 2010.

[Ramos-Llordén 16] Gabriel Ramos-Llordén, J Arnold & Jan Sijbers. *Partial discreteness: A novel prior for magnetic resonance image reconstruction.* IEEE transactions on medical imaging, vol. 36, no. 5, pages 1041–1053, 2016.

[Ramos Llordén 18] Gabriel Ramos Llordén. *Improved MRI relaxometry through statistical signal processing.* PhD thesis, University of Antwerp, 2018.

[Rangan 16] Sundeep Rangan, Philip Schniter, Erwin Riegler, Alyson K Fletcher & Volkan Cevher. *Fixed points of generalized approximate message passing with arbitrary matrices.* IEEE Transactions on Information Theory, vol. 62, no. 12, pages 7464–7474, 2016.

[Rangan 19] Sundeep Rangan, Philip Schniter & Alyson K Fletcher. *Vector approximate message passing.* IEEE Transactions on Information Theory, 2019.

[Rangana 19] Sundeep Rangana, Philip Schniterb, Alyson K Fletcherc & Subrata Sarkar. *On the convergence of approximate message passing with arbitrary matrices.* IEEE Transactions on Information Theory, 2019.

[Rasche 99] Volker Rasche, Roland Proksa, R Sinkus, Peter Bornert & Holger Eggers. *Resampling of data between arbitrary grids using convolution interpolation.* IEEE transactions on medical imaging, vol. 18, no. 5, pages 385–392, 1999.

[Ravishankar 11] Saiprasad Ravishankar & Yoram Bresler. *MR image reconstruction from highly undersampled k-space data by dictionary learning.* IEEE Trans. Med. Imag, vol. 30, no. 5, pages 1028–1041, 2011.

[Roemer 90] Peter B Roemer, William A Edelstein, Cecil E Hayes, Steven P Souza & Otward M Mueller. *The NMR phased array.* Magnetic resonance in medicine, vol. 16, no. 2, pages 192–225, 1990.

[Roman 14] Bogdan Roman, Anders Hansen & Ben Adcock. *On asymptotic structure in compressed sensing.* arXiv preprint arXiv:1406.4178, 2014.

[Rudin 92] Leonid I Rudin, Stanley Osher & Emad Fatemi. *Nonlinear total variation based noise removal algorithms.* Physica D: Nonlinear Phenomena, vol. 60, no. 1, pages 259–268, 1992.

[Samsonov 04] Alexei A Samsonov, Eugene G Kholmovski, Dennis L Parker & Chris R Johnson. *POCSENSE: POCS-based reconstruction for sensitivity encoded magnetic resonance imaging.* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, vol. 52, no. 6, pages 1397–1406, 2004.

[Samsonov 10] Alexey A Samsonov, Julia Velikina, Youngkyoo Jung, Eugene G Kholmovski, Chris R Johnson & Walter F Block. *POCS-enhanced correction of motion artifacts in parallel MRI.* Magnetic resonance in medicine, vol. 63, no. 4, pages 1104–1110, 2010.

[Setzer 09] Simon Setzer. *Split Bregman algorithm, Douglas-Rachford splitting and frame shrinkage.* In International Conference on Scale Space and Variational Methods in Computer Vision, pages 464–476. Springer, 2009.

[Shannon 49] Claude Elwood Shannon. *Communication in the presence of noise.* Proceedings of the IRE, vol. 37, no. 1, pages 10–21, 1949.

[Shapiro 93] Jerome M Shapiro. *An embedded hierarchical image coder using zerotrees of wavelet coefficients.* In 1993 Data Compression Conference, pages 214–223. IEEE, 1993.

[Shensa 92] Mark J Shensa. *The discrete wavelet transform: wedding the a trous and Mallat algorithms.* IEEE Transactions on signal processing, vol. 40, no. 10, pages 2464–2482, 1992.

[Simoncelli 97] Eero P Simoncelli. *Statistical models for images: Compression, restoration and synthesis.* In Signals, Systems & Computers, 1997. Conference Record of the Thirty-First Asilomar Conference on, volume 1, pages 673–678. IEEE, 1997.

[Skodras 01] Athanassios N Skodras, Charilaos A Christopoulos & Touradj Ebrahimi. *JPEG2000: The upcoming still image compression standard.* Pattern Recognition Letters, vol. 22, no. 12, pages 1337–1345, 2001.

[Smith 13] David S Smith, Lori R Arlinghaus, Thomas E Yankeelov & Edward B Welch. *Curvelets as a sparse basis for compressed sensing magnetic resonance imaging.* In SPIE Medical Imaging, pages 866929–866929. International Society for Optics and Photonics, 2013.

[Som 12] Subhojit Som & Philip Schniter. *Compressive imaging using approximate message passing and a Markov-tree prior.* IEEE transactions on signal processing, vol. 60, no. 7, pages 3439–3448, 2012.

[Sonka 14] Milan Sonka, Vaclav Hlavac & Roger Boyle. Image processing, analysis, and machine vision. Cengage Learning, 2014.

[Starck 02] Jean-Luc Starck, Emmanuel J Candès & David L Donoho. *The curvelet transform for image denoising.* IEEE Trans. Image Process., vol. 11, no. 6, pages 670–684, 2002.

[Starck 04] Jean-Luc Starck, DL Donoho & Michael Elad. *Redundant multiscale transforms and their application for morphological component separation.* Rapport technique, CM-P00052061, 2004.

[Starck 05] Jean-Luc Starck, Michael Elad & David L. Donoho. *Image Decomposition via the Combination of Sparse*

*Representations and a Variational Approach*. IEEE Trans. Image Process., vol. 14, no. 10, pages 1570–1582, 2005.

[Starck 07] Jean-Luc Starck, Jalal Fadili & Fionn Murtagh. *The undecimated wavelet decomposition and its reconstruction*. IEEE Transactions on Image Processing, vol. 16, no. 2, pages 297–309, 2007.

[Sutton 03] Bradley P Sutton, Douglas C Noll & Jeffrey A Fessler. *Fast, iterative image reconstruction for MRI in the presence of field inhomogeneities*. IEEE transactions on medical imaging, vol. 22, no. 2, pages 178–188, 2003.

[Tessens 08] Linda Tessens, Aleksandra Pizurica, Alin Alecu, Adrian Munteanu & Wilfried R Philips. *Context adaptive image denoising through modeling of curvelet domain statistics*. Journal of Electronic Imaging, vol. 17, no. 3, page 033021, 2008.

[Tsai 00] Chi-Ming Tsai & Dwight G Nishimura. *Reduced aliasing artifacts using variable-density k-space sampling trajectories*. Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, vol. 43, no. 3, pages 452–458, 2000.

[Tsao 03] Jeffrey Tsao, Peter Boesiger & Klaas P Pruessmann. *k-t BLAST and k-t SENSE: dynamic MRI with high frame rate exploiting spatiotemporal correlations*. Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, vol. 50, no. 5, pages 1031–1042, 2003.

[Uecker 14] Martin Uecker, Peng Lai, Mark J Murphy, Patrick Virtue, Michael Elad, John M Pauly, Shreyas S Vasanawala & Michael Lustig. *ESPIRiT an eigenvalue approach to autocalibrating parallel MRI: where SENSE meets GRAPPA*. Magnetic resonance in medicine, vol. 71, no. 3, pages 990–1001, 2014.

[Vila 15] Jeremy Vila, Philip Schniter, Sundeep Rangan, Florent Krzakala & Lenka Zdeborová. *Adaptive damping and mean removal for the generalized approximate message passing algorithm.* In 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 2021–2025. IEEE, 2015.

[Walsh 00] David O Walsh, Arthur F Gmitro & Michael W Marcellin. *Adaptive reconstruction of phased array MR imagery.* Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, vol. 43, no. 5, pages 682–690, 2000.

[Wang 04] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh & Eero P Simoncelli. *Image quality assessment: from error visibility to structural similarity.* IEEE Trans. Image Process., vol. 13, no. 4, pages 600–612, 2004.

[Wang 08] Yilun Wang, Junfeng Yang, Wotao Yin & Yin Zhang. *A new alternating minimization algorithm for total variation image reconstruction.* SIAM Journal on Imaging Sciences, vol. 1, no. 3, pages 248–272, 2008.

[Wang 10] Xiang-Yang Wang, Hong-Ying Yang & Zhong-Kai Fu. *A new wavelet-based image denoising using undecimated discrete wavelet transform and least squares support vector machine.* Expert Systems with Applications, vol. 37, no. 10, pages 7040–7049, 2010.

[Whittaker 15] Edmund Taylor Whittaker. *On the functions which are represented by the expansions of the interpolation-theory.* Proceedings of the Royal Society of Edinburgh, vol. 35, pages 181–194, 1915.

[Winkelmann 07] Stefanie Winkelmann, Tobias Schaeffter, Thomas Koehler, Holger Eggers & Olaf Doessel. *An optimal radial profile order based on the Golden Ratio*

*for time-resolved MRI.* IEEE Trans. Med. Imag, vol. 26, no. 1, pages 68–76, 2007.

[Winkler 12] Gerhard Winkler. Image analysis, random fields and markov chain monte carlo methods: a mathematical introduction, volume 27. Springer Science & Business Media, 2012.

[Woods 11] John W Woods. Multidimensional signal, image, and video processing and coding. Academic press, 2011.

[Wright 97] G Wright. *Magnetic Resonance Imaging.* IEEE Signal Process. Mag., vol. 14, no. 1, pages 56–66, 1997.

[Wright 09] Stephen J Wright, Robert D Nowak & Mário AT Figueiredo. *Sparse reconstruction by separable approximation.* IEEE Trans. Signal Process., vol. 57, no. 7, pages 2479–2493, 2009.

[Wright 14] Katherine L Wright, Jesse I Hamilton, Mark A Griswold, Vikas Gulani & Nicole Seiberlich. *Non-Cartesian parallel imaging reconstruction.* Journal of Magnetic Resonance Imaging, vol. 40, no. 5, pages 1022–1040, 2014.

[Xu 07] Jinjun Xu & Stanley Osher. *Iterative regularization and nonlinear inverse scale space applied to wavelet-based denoising.* IEEE Transactions on Image Processing, vol. 16, no. 2, pages 534–544, 2007.

[Yang 10] Junfeng Yang, Yin Zhang & Wotao Yin. *A fast alternating direction method for $TV\ell_1$-$\ell_2$ signal reconstruction from partial Fourier data.* Selected Topics in Signal Processing, IEEE Journal of, vol. 4, no. 2, pages 288–297, 2010.

[Yen 00] Gary G Yen & K-C Lin. *Wavelet packet feature extraction for vibration monitoring.* IEEE transactions on industrial electronics, vol. 47, no. 3, pages 650–667, 2000.

[Yin 08] Wotao Yin, Stanley Osher, Donald Goldfarb & Jerome Darbon. *Bregman iterative algorithms for*

*$\ell_1$-minimization with applications to compressed sensing*. SIAM Journal on Imaging Sciences, vol. 1, no. 1, pages 143–168, 2008.

[Ying 04] Leslie Ying, Dan Xu & Z-P Liang. *On Tikhonov regularization for image reconstruction in parallel MRI*. In The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, volume 1, pages 1056–1059. IEEE, 2004.