Technische Universität München

Zentrum Mathematik

Acceleration of the Spatial Selective Excitation of MRI via Sparse Approximation

Dong Chen

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaft (Dr.rer.nat.)

genehmigten Dissertation.

Vorsitzender :
Prüfer der Dissertation :
1. Univ.-Prof. Dr. Folkmar Bornemann
2. Univ.-Prof. Dr. Peter Rentrop
3. Prof. Dr. Yudong Zhu
— New York University, USA (schriftliche Beurteilung)

Die Dissertation wurde am 15.07.2009 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 24.11.2009 angenommen.

To my wife Xiaohong, our Kuan
and my parents.

Acknowledgements

This thesis would not be possible without the help and support I got from many sides.

At first I would like to express my deep thanks to Prof. Bornemann. I first got to know Prof. Bornemann through his four semester lecture series of "Numerical Mathematics" and the subsequent lecture of "PDE in Image Processing". They are by far the best lectures I've ever had. As my PhD advisor he always provides constructive suggestions and valuable insight to the underlying mathematical problems. At the same time, he encourages independent work and grants sufficient freedom for trying and developing my own idea. The most valuable thing I learned from him is how to think and approach a problem.

The next I want to thank GE Global Research for supporting this thesis. My daily work is at Imaging Technology Lab, GE Global Research Center, Garching. This work is entirely embedded into the parallel transmit project in the lab financed by GE Healthcare. I want to thank the project supervisors I worked with during this period. I learned a lot from them — their profound MRI knowledge, project-management skills and social competence. They are Dr. Mika Vogel (Jan. 08 – Now), Dr. Yudong Zhu (Jan.08 – Okt. 09), Dr. Hans-Peter Fautz (May. 07 – Dec. 08). Yudong also keeps providing me helpful discussions and suggestions even after his leaving GE for his academic career in NYU.

I want to thank Jonathan Sperl, Johannes Haas and Silke Lechner for reading different parts of the thesis and provided very helpful suggestions.

At last but most important, I want to thank my family, my wife Xiaohong, our son Kuan and my parents, for all the support Ive received.

Contents

Symbol overview

The abbreviations:

Symbol	Explanation
(D)FT/(F)FT	(Discrete) Fourier Transform/(Fast) Fourier Transform
FOV	Field Of View
MR(I)	Magnetic Resonance (Imaging)
NMR	Nuclear Magnetic Resonance
pTx	Parallel excitation/Parallel RF transmit
pRx	Parallel receive/Parallel imaging
R	Reduction factor
RF	Radio Frequency
RRMS(E)	Relative Root Mean Squares (Error)
SLI	Subspace Linear Independent
SNR	Signal Noise Ratio
SSEP	Spatial Selective Excitation Problem
STA	Small-Tip-Angle
TOL, or ε	User tolerance

The symboles

Symbol	Explanation
$\vec{B}$	The magnetic field. $\vec{B} : (\xi, t) \mapsto \vec{M}(\xi, t) \in \mathbb{R}^3$
B_{xy}	The xy field components. $B_{xy} : (\xi, t) \mapsto B_{xy}(\xi, t) \in \mathbb{C}$
$\boldsymbol{B}, \boldsymbol{b} \in \mathbb{C}^M$ (/ **RF**)	The discrete RF waveform as a vector.
$C \in \mathbb{N}, \sim \mathcal{O}(10)$	the number of the parallel transmit channels (coils)
$c \in \{1, \cdots, C\}$	The index of the channels (coils)
$\mathrm{Dom}(\mathcal{A})$	Domain/row-space of matrix $\mathcal{A}$,
$\mathcal{E}$, or $\mathcal{E}_{\boldsymbol{k}} \in \mathbb{C}^{m \times n}$	The encoding matrix using k sampling locations $\boldsymbol{k}$
$\hat{\mathcal{E}}^{(c)}$/ $\hat{\mathcal{E}}^{(c)}_{\boldsymbol{k}} \in \mathbb{C}^{n \times m}$	$\hat{\mathcal{E}}^{(c)}_{\boldsymbol{k}} := \mathcal{E}_{\boldsymbol{k}} \cdot \mathcal{S}^{(s)}$, the sensitivity including encoding matrix
ε (or TOL)	User tolerance
$\boldsymbol{f}^{(k)} \in \mathbb{C}^N$	the Fourier harmonics vector in spatial domain, $\boldsymbol{f}^{(k)}_j = e^{i \cdot k \cdot \xi_j}$
$\hat{\boldsymbol{f}}^{(k,c)} \in \mathbb{C}^N$	$\hat{\boldsymbol{f}}^{(k,c)} := \mathcal{S}^{(c)} \cdot \boldsymbol{f}^{(k)}$
g	The gradient waveform. $g : t \mapsto g(t) \in \mathbb{C}$
$\langle g(t), \xi \rangle$	For notation simplicity, we abuse the $\langle \cdot, \cdot \rangle$ notation: $\langle g(t), \xi \rangle := \mathrm{real}(\xi) \cdot \mathrm{real}(g(t)) + \mathrm{imag}(\xi) \cdot \mathrm{imag}(g(t))$
k	The continuous feasible k space trajectory. $k : t \mapsto k(t) \in \mathbb{C}$
$\boldsymbol{k} \in \mathbb{C}^M$	The discrete form of the k, with $4\mu s$ sampling distant.
$\boldsymbol{\kappa} \in \mathbb{C}^L$	The sparse obligatory sampling locations to be covered by k.
L or l	#(participating vectors) or #(sparse sampling locations)
Λ	The index vector for the selected sampling location with respect to a large set of candidate k sampling locations.
$\lvert\Lambda\rvert$	The number of the element of the index vector *Lambda*
$M, m \in \mathbb{N},$	The dimension of the discrete frequency domain (k space).
$\vec{M}$	The 3D magnetization vector. $\vec{M} : (\xi, t) \mapsto \vec{M}(\xi, t) \in \mathbb{R}^3$
M_{xy}	The xy magnetization. $M_{xy} : (\xi, t) \mapsto M_{xy}(\xi, t) \in \mathbb{C}$
$\boldsymbol{M} \in \mathbb{C}^N$	The discrete xy-magnetization profile as a vector.

to be continued in next page

Symbol	Explanation
$N, n \in \mathbb{N}$	The dimension of the discrete spatial domain, #(pixels).
$\mathbb{O} \subset \mathbb{C}$	The spatial region of the object.
$\Omega \in \mathbb{N}$	The number of the atoms, dictionary cardinality.
$\boldsymbol{P}, \boldsymbol{p} \in \mathbb{C}^N$	The target vector for the sparse approximation.
$\mathfrak{P}$, $\mathfrak{p}$	The target vector space.
Range($\mathcal{A}$)	Range/column-space of matrix $\mathcal{A}$.
RF	The RF waveform. RF : $t \mapsto \mathrm{RF}(t) \in \mathbb{C}$
$\mathbf{RF} \in \mathbb{C}^M$ (or $\boldsymbol{B}$)	Discrete RF waveform as a vector
S	The detected signal in the imaging problem. $S : t \mapsto S(t) \in \mathbb{C}$
$\boldsymbol{S} \in \mathbb{C}^M$	The discrete signal in the imaging problem.
$\mathfrak{S}$	The signal space.
s	The spatial RF sensitivity profile. $s : \xi \mapsto s(\xi) \in \mathbb{C}$
$\boldsymbol{s} \in \mathbb{C}^N$	The discrete form of the sensitivity profile as vector.
$\mathcal{S} \in \mathbb{C}^{N \times N}$	$\mathcal{S} := \mathrm{Diag}(\boldsymbol{s})$. The sensitivity matrix.
$\mathfrak{s}$	The subspace atom.
$\mathfrak{s}_\Lambda$	The set of subspace atoms. $\mathfrak{s}_\Lambda := \{\mathfrak{s}_{\Lambda(1)}, ..., \mathfrak{s}_{\Lambda(end)}\}$
$\mathcal{T}$, or $\mathcal{T}_{\boldsymbol{k}} \in \mathbb{C}^{N \times M}$	The transmit matrix using the sampling locations $\boldsymbol{k}$, (4.11)
$\hat{\mathcal{T}}_{\boldsymbol{k}}^{(c)} \in \mathbb{C}^{N \times M}$	$\hat{\mathcal{T}}_{\boldsymbol{k}}^{(c)} := \mathcal{S}^{(c)} \cdot \mathcal{T}_{\boldsymbol{k}}$, The sensitivity including transmit matrix.
$\xi \in \mathbb{C}$	The spatial coordinate (pixel location).
$\delta\xi \in \mathbb{R}$	The pixel size.
$\overset{\mathrm{TOL}}{=}$ or $\overset{\epsilon}{=}$	Equalty in the sense of least squares regarding TOL or ϵ.

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The Problem and Result Overview

This work is an applied mathematics thesis with Magnetic Resonance Imaging (MRI) as application. It is an interdisciplinary work of applied mathematics and medical imaging.

Magnetic resonance imaging is a rapidly developed medical imaging technology. Compared to the other medical imaging modalities, it is distinguished by its non-ionizing-radiation nature and the greater contrast between different soft tissues of the body.

The functioning principle of MRI can be divided into two stages.

1. the *excitation* stage:
   In order to be detected at all, the spins of the body need to be excited at first.

2. the *imaging* stage:
   The excited spins of the body emit electro-magnetic waveform, which will be detected by the receive coil. The detected signal will be used to reconstruct the image of the body.

An important task of the excitation stage is the so-called *spatial selective excitation*. The goal is to control the spatially distributed spin ensemble to a given spatial profile. One calls this given spatial profile the *target (excitation) profile*.

There are two major reasons for a *fast* spatial selective excitation process.

1. The model of relaxation-free Bloch equation neglects the relaxation process. The use of this simplified model leads to the restriction that

all the solutions based on it are only valid when the excitation duration is small enough regarding the spin *relaxation time*. The solutions from the current conventional method have usually such a long excitation duration that it is not possible to perform a real 3D spatial selective excitation, which is relevant for the praxis application.

2. In addition, the hardware realization accuracy is usually limited. By very long excitation duration, the hardware imperfection could accumulate to serious excitation artifact at the end. This issue is especially problematic by the high field scanner like 3T or higher.

These two major reasons motivate the acceleration of the spatial selective excitation problem. (In contrast, the motivation to accelerate the imaging stage is to reduce the patient scanning duration.)

Physically, the excitation process is controlled by the external magnetic field, $\vec{B} : (\xi, t) \mapsto \vec{B}(\xi, t) \in \mathbb{R}^3$, where $\xi \in \mathbb{O}$ is the spatial coordinate, $t \in [0, t_f]$ is the time. The $\vec{B}$ is realized by two hardware controls, the *gradient* and the *RF*:

1. The so-called *gradient coil* associates with the external magnetic field in z-direction, $B_z : (\xi, t) \mapsto B_z(\xi, t) \in \mathbb{R}$.

   The B_z consists of two parts: $B_z(\xi, t) = B_0 + \langle \xi, g(t) \rangle$

   (a) The spatially and temporally constant component (e.g. 1.5T or 3T etc), one calls it B_0.

   (b) The spatially linearly varying part. Its spatial slope is called the *gradient*.

   The gradient coil controls the spatial slope of B_z. The gradient as a function of time is called the gradient waveform $g : [0, t_f] \mapsto g(t) \in \mathbb{C}$.

2. The so-called *RF transmit coil* controls the external magnetic field in xy-direction, $RF : t \mapsto RF(t) \in \mathbb{C}$. The spatial local xy magnetic field $B_{xy}(\xi, t)$ is the global $RF(t)$ multiplied by a complex valued spatial sensitivity factor $s : \xi \mapsto s(\xi)\mathbb{C}$. The spatial function s is called the *sensitivity profile*.

$$B_{xy}(\xi, t) = s(\xi) \cdot RF(t)$$

From a physical aspect, the RF waveform control is responsible for that the spin get flipped (excited) from the ground state $(0, 0, 1)^T$ at all. The gradient waveform control is responsible for the capability to distinguish the

spins from spatial location to spatial location (the voxels), so that the spatial selectivity would be possible.

The MRI scanner components, the positioning of the gradient coil and the RF coil are shown in the fig 1.1.

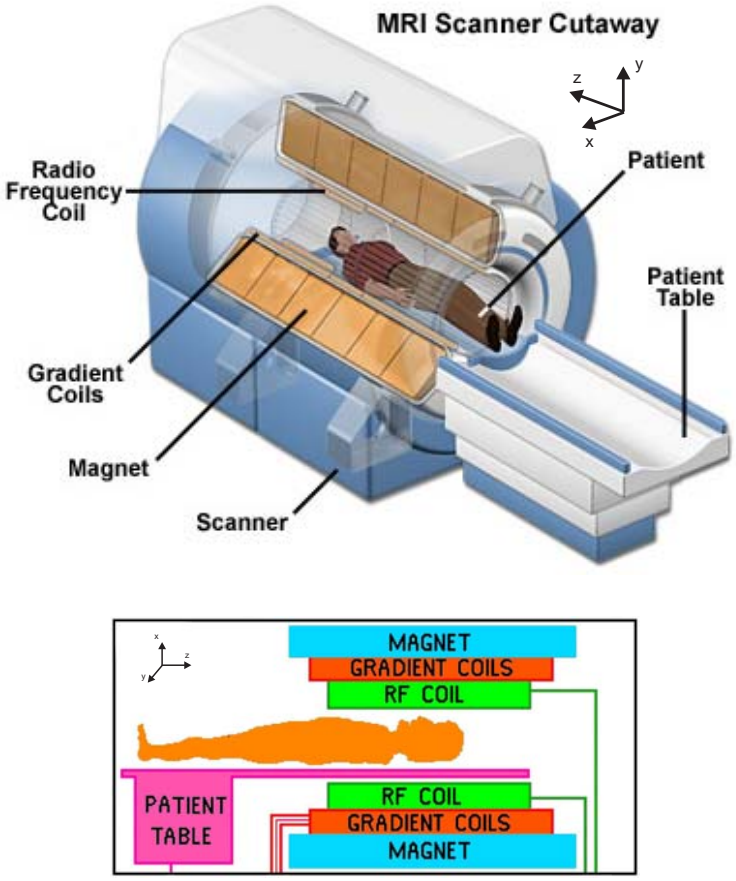


Figure 1.1: MRI scanner schematic

In a standard MRI scanner, there are a single Gradient control and a single RF transmit control. An extension of the standard setting is the so-called *parallel RF transmit*. By parallel RF transmit, instead of a single RF transmit control (coil), there are a series of independent RF transmit controls (coils). They are also called the *RF transmit channels*. Each RF transmit coil has different spatial sensitivity profile. All transmit channels share the same gradient control. Thank to the additional degree of freedom provided by the parallel RF control channels, one gets the capability to shorten the excitation duration.

This extension from the standard single channel RF transmit to multi-

channel parallel RF transmit is schematically demonstrated in the fig 1.2. The acceleration effect of the parallel transmit setting can be seen by the comparison of the column A and column B in the fig 1.4.

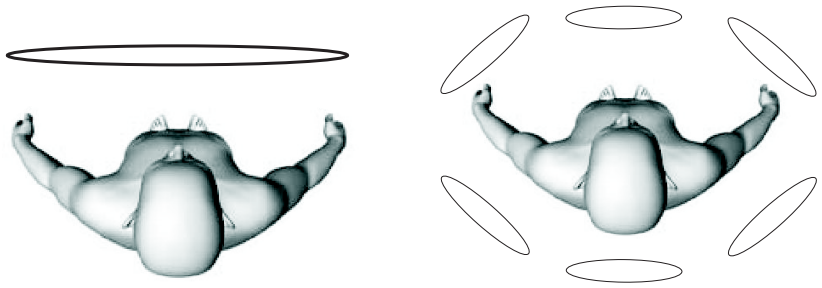


Figure 1.2: Single channel RF transmit and Parallel RF transmit

The conventional paradigm to design the $g(t)$ and $RF(t)$ for the spatial selective excitation problem is based on the so-called STA model [PNM89a]. In the STA model the spatial selective excitation problem can be transferred to a linear representation problem using the Fourier vectors, $\boldsymbol{f}_j^k = e^{ik\xi_j}$. The discretely sampled gradient waveform control is associated with the frequency parameter of the Fourier vectors k. There is a bijective mapping between g and k,

$$k(t) := \int_{t_f}^{t} g(\tau) d\tau.$$

The discretely sampled RF waveform is the representation coefficients of the according Fourier vectors.

In the conventional paradigm the frequency parameter k is determined by the Shannon-Nyquist theorem. The discrete values of the RF waveform are the according Fourier coefficients of the Nyquist sampling locations.

This paradigm is demonstrated in fig 1.3. One starts with the target profile 1). From the spatial sampling grid of the target profile one determines the according Nyquist grid (2). Using both (1) and (2) one calculates the RF waveform as the coefficients of the Fourier vectors (3). The gradient waveform is calculated as the derivative of the k space sampling trajectory, (2) $\rightarrow$ (4). At the end the two control waveforms, RF and gradient, are validated by the Bloch equation simulation (5) and the phantom experiment on the MRI scanner (6).

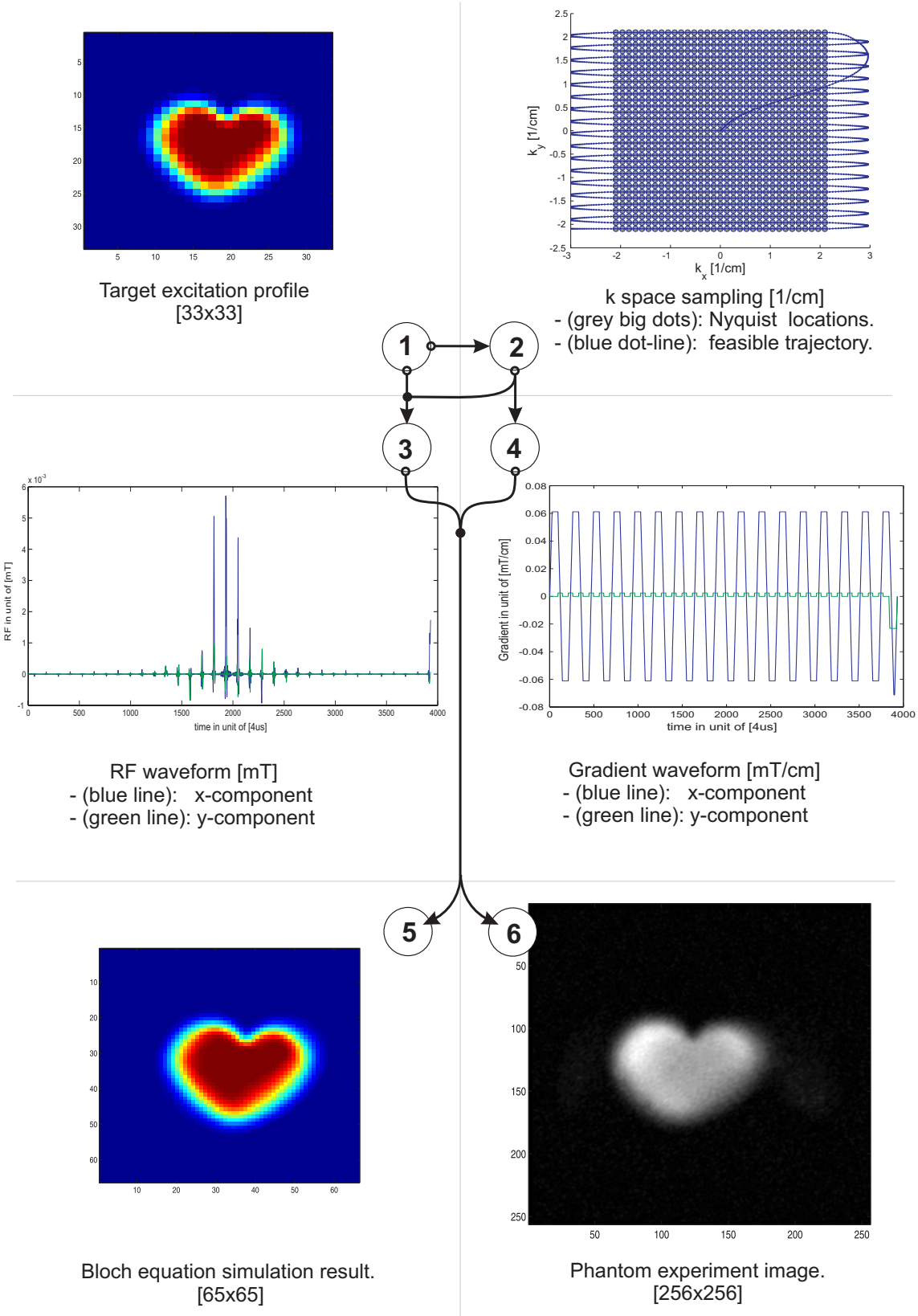


Figure 1.3: The conventional paradigma of full Nyquist STA method

At first sight, this paradigm provides for any target excitation profile a successful excitation control. However, a problem of this paradigm is that it fixedly requires covering all the Nyquist sampling locations, which results a very long excitation duration. By 2D excitation, it makes the resulting excitation profile very vulnerable to the hardware imperfection. The 3D excitation is almost impossible due to the relaxation time.

However, the Shannon-Nyquist theorem is a sufficient but not a necessary condition for the successful representation of a target vector with the the Fourier vectors. We propose an adaptive concept to determine a sparse obligatory sampling locations instead of just completely covering all the Nyquist sampling locations in k space.

The realization of this adaptive concept leads to the mathematical topic of *sparse approximation*. The topic will be discussed in Chapter-5.

Due to the gradient hardware feasibility, an additional step is necessary. We call this step the fast covering $\boldsymbol{\kappa}$ problem. The goal is to fast cover the selected obligatory locations $\boldsymbol{\kappa} \in \mathbb{C}^L$ in k space by a continuous traveling trajectory with constraints on the traveling velocity and the traveling acceleration. More details regarding this step is discussed in §-7.4.

At the end, the proposed adaptive paradigm can deliver solutions with much shorter excitation duration than from the conventional paradigm. (Compare column-A with column-C in fig 1.4.) More details on the validation result of the acceleration effect are discussed in §-7.7 and §-7.8.

Our second goal is to combine the adaptive method with the *parallel RF transmit* setting to maximize the accelerating capability.

The combination of the adaptive concept with the parallel transmit setting brings additional difficulty. By the single channel setting, each k sampling location associates with a single Fourier vector. In the parallel transmit setting, due to the fact that the gradient is "shared" by the multiple RF transmit channels, each k space sampling location associates with several modified Fourier vectors. Minimizing the number of k sampling locations leads to the special sparsity requirement that not the number of the Fourier vectors should be minimized, but the number of the groups of the vectors associating with the same k locations should be minimized. We call this special sparsity requirement the *synchronized sparsity*.

The problem of sparsifying the k space sampling can then be formulated as a modified version of the conventional sparse approximation problem. We call it the synchronized sparse approximation or subspace approximation (Chapter-6). Accordingly, the numerical method is developed as a general-

ized version of the conventional *OMP method*. We call it subspace-OMP. The theoretical analysis of the subspace sparse approximation and subspace-OMP method are also provided.

With the mathematical tool of subspace sparse approximation, we are able to successfully integrate the adaptive concept to the parallel RF transmit setting. One compares the column-B with the column-D,E in fig 1.4. (More details on the combining the adaptive paradigm with the parallel transmit setting are discussed in Chapter-8.)

We systematically validated the proposed adaptive paradigm in three stages. (§-7.5, §-8.3)

I. The validation of mathematic tools — the algorithms for (subspace) sparse approximation (§-7.6, §-8.4).

II. The validation of the concept via the simulation of the Bloch equation (§-7.7, §-8.5). A Bloch equation simulation tool are developed to evaluate the method per simulation.

III. The validation of the practical applicability via the real-world phantom experiment on the MRI scanner (§-7.8, §-8.6).

Fig 1.4 is a representative overview of the core results of this work. Top-left is the excitation error vs. duration plot. Usually we don't really care about the errors below 5%, because the hardware realization imperfection is already about 5%. The column A is the conventional non adaptive paradigm with single transmit channel setting. The column B is the conventional non adaptive paradigm using a six-channel parallel transmit setting. The column C is the adaptive paradigm with single channel setting. The column D and E are the adaptive parallel transmit approach with two difference accuracy levels. One sees clearly the acceleration benefit of the adaptive solutions.

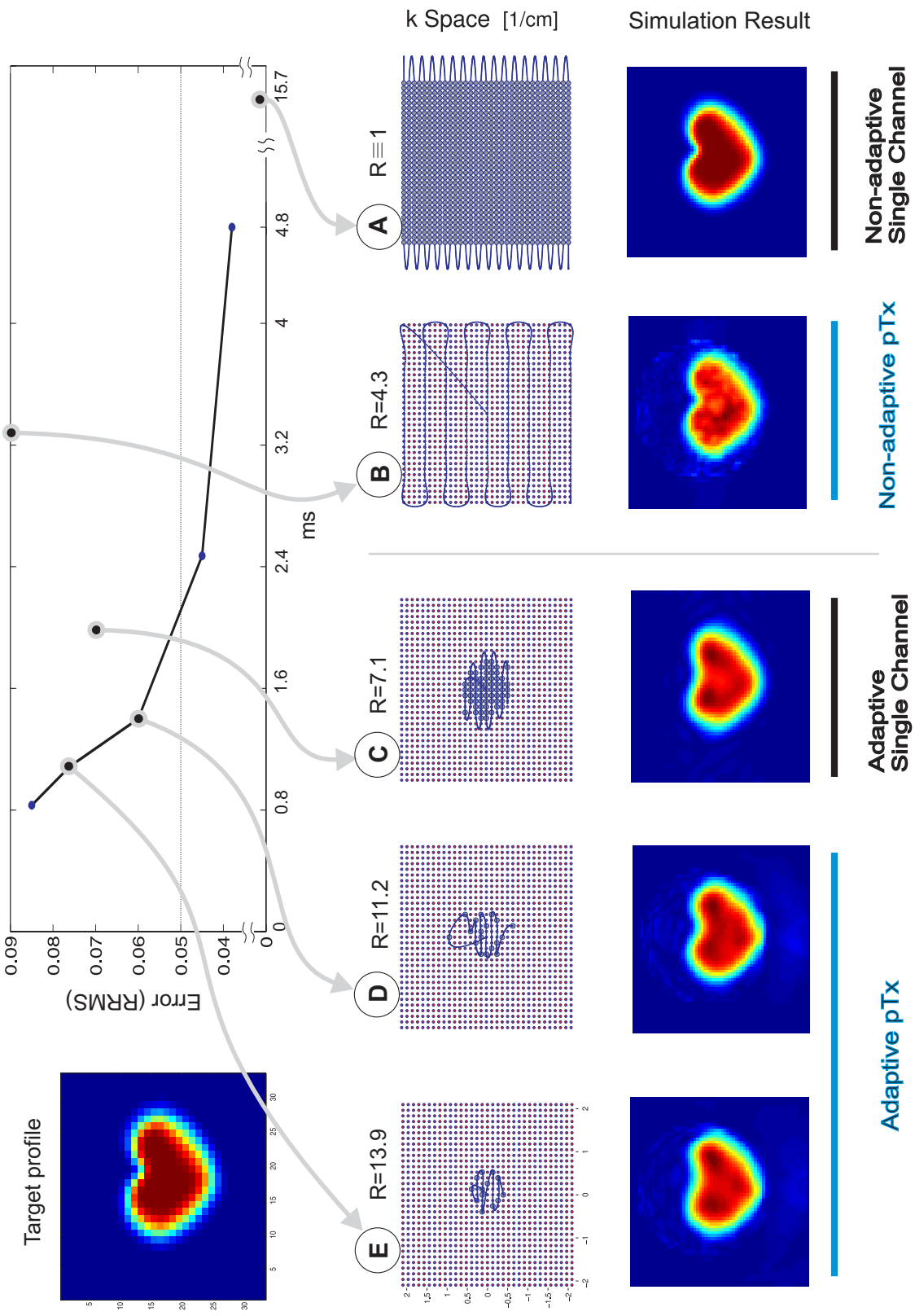


Figure 1.4: The results overview

1.2 The Outline of the Thesis

The thesis is divided into three parts according to their roles in the interdisciplinary nature of this work:

In the first part, we transfer from the MRI application problems to the according mathematical models and problems. We will go through the two strongly related problems, 1) the acceleration of the *imaging stage* and 2) the acceleration of the *excitation stage*. Our application focus is on the acceleration of the excitation process. However, since a) the *imaging process* and the *excitation process* have symmetric structure, b) the imaging process and its acceleration have been more intensively investigated in the MRI community than the excitation process, it is helpful to put both problems and their existing methods together in a general common frame. For that purpose, we will go through the major acceleration methods for both *imaging* and *excitation* processes: the *parallel imaging* method and the *compressed sensing* method from the imaging side and the *parallel RF transmit* method from the excitation side. After putting them in a common mathematical frame, the adaptive concept and its the connection with the sparse approximation will appear straightforwardly.

The second part is the mathematical part. The goal of this part is to develop the necessary mathematical tools to realize the adaptive concept, especially the synchronized sparsity requirement, which is needed by combining the adaptive approach with the parallel transmit setup. We will introduce the sparse approximation topic and its state of the art methods. We will then generalize the original problem statement to the subspace sparse approximation problem for synchronized sparsity. The standard *Orthogonal Matching Pursuit(OMP)* method will be generalized to the so-called *Subspace-OMP* to approach the new problem. The according theoretical analysis will also be extended.

In the third part we transfer from the mathematic tools back to the MRI application. The developed method is implemented in a Matlab software package. The validation of the proposed method is performed in two levels: Computer simulation of the Bloch equation and the real world MRI scanner phantom experiment on a GE 3T MRI scanner.

The realization of the adaptive concept in the application is divided into two parts according to the two hardware setups. The first part is the implementation and validation of the adaptive method without interfering with the existing accelerating method of *parallel RF transmits*. This part will be

introduced in Chapter-7. The second part, and the final target, is to embed the adaptive concept into the existing *parallel RF transmit* setting and thus maximally exhaust the accelerating capability. This part of the work will be introduced in the chapter-8.

We follow the spirit of the reproducible research [DMR+08]. The Matlab codes are available as a CD-Rom for the verification and re-usage purpose. After a short 'Getting-Started Guide', the user can easily reproduce the results in this work.

Part I

The Transfer from MRI to Sparse Approximation

Chapter 2

Basic Principles of MRI

2.1 The Motion of Nuclear Magnetization

The nuclear spin is a quantum mechanical property. However, the typical scale in the application of medical imaging is [mm]. At the scale of [mm] or larger, one can sufficiently describe collective behavior of the spins macroscopically as the magnetization motion in the classical picture.

The macroscopic equation of motion for a collective behavior of the nuclear magnetization is the Bloch equation [Blo46][Jay55]. Let's assume the initial equilibrium magnetization state is in the z direction $\vec{e}_z$, then the Bloch equation is

$$\dot{\vec{M}}(t) = \vec{M}(t) \times \gamma \cdot \vec{B}(t) + \frac{M_z^0 - M_z(t)}{T_1}\vec{e}_z + \frac{M_x(t)\vec{e}_x + M_y(t)\vec{e}_y}{T_2}. \qquad (2.1)$$

Where:

- $\vec{e}_{x,y,z}$ are the canonical basis in $\mathbb{R}^3$.
- $\vec{M} \equiv (M_x, M_y, M_z)^T : t \mapsto \vec{M}(t) \in \mathbb{R}^3$ is the magnetization vector.
- $\vec{B} : t \mapsto \vec{B}(t) \in \mathbb{R}^3$ is the vector of the external magnetic field.
- '$\times$' is the vector cross-product.
- M_z^0 is the z-direction magnetization at the initial equilibrium state. Without loss of generality, we can assume $M_z^0 = 1$.
- T_1, T_2 are constants characterizing the relaxation process (the relaxation time).

– γ is constant called gyromagnetic ratio. γ is of the unit radian per second per Tesla: $(s^{-1} \cdot T^{-1})$.

Physically the dynamics of the nuclear magnetization consists of two types of motions: *precession* and *relaxation*.

1. *Precession:* (the first term to the right side of the Bloch equation (2.1))

   In an external magnetic field, $\vec{B}$, the magnetization vector $\vec{M}$ will precess around $\vec{B}$. The precession frequency ω_0 is proportional to the magnitude of the external field $|\vec{B}|$ and the angle $\alpha = \angle(\vec{B}, \vec{M})$

   $$\omega_0(t) = \gamma \cdot \sin(\alpha(t)) \cdot |\vec{B}(t)|$$

   One calls this frequency ω_0 the Larmor frequency. With other words, whenever $\vec{M} \nparallel \vec{B}$, one will observe the precession motion around $\vec{B}$ at the Larmor frequency.

   We know that any accelerating charges or magnetic dipoles will induce electro-magnetic waveform. Hence the procession motion will always emit an electro-magnetic waveform with a frequency equals the Larmor frequency.

2. *Relaxation:* (the 2nd and 3rd term to the right side the Bloch equation (2.1))

   The precession process is accompanied by the relaxation process, which is a process wherein the magnetization vector recovers towards the lowest state of energy $\hat{e}_z$.

   The relaxation process can be phenomenologically described by the recovery process in two orthogonal directions. i) The M_z component recovers towards value 1. This is described by the phenomenological parameter T_1 (2nd term right side the Bloch equation). ii) The $|M_{\text{xy}}|$ component recovers towards value 0. This is described by the phenomenological parameter T_2 (3rd term to the right side of the Bloch equation (2.1)).

   Physically, the relaxation process can be divided into two types according to their physical properties:

   i. Due to energy conservation, by emitting electro-magnetic waveform and thus losing energy, the excited magnetization ($\alpha > 0$) will recover towards the energy ground state ($\alpha = 0$).

      In this type of relaxation, the magnetization vector exhibits norm conservation.

The recovery process in M_z is caused solely by this energy conservation reason. In MR terminology, one calls this type of relaxation the T1 relaxation.

ii. In addition, the $|M_{xy}|$ component will decay faster due to the phase-incoherency of M_{xy} among the nuclear spins within an individual pixels (the dephasing effect). Collectively, one observes a faster reduction of the $|M_{xy}|$ than $\sqrt{1-M_z^2}$. In MR terminology one calls this type of relaxation the T2 relaxation.

This type of additional $|M_{xy}|$ decay is responsible for the *non* norm-conservation behavior of the magnetization vector during the relaxation process.

The dephasing mechanism has no effect on the M_z recovery process and hence it has no influence on T_1. That's why T_2 is usually much smaller than T_1.

2.2 Nuclear Magnetic Resonance (NMR) Principle

By NMR one has certain object, e.g. certain complex material. The goal is to obtain certain information about the microscopic properties of this object. The basic concept doing that via the magnetic resonance principle can be demonstrated in fig.-2.1.

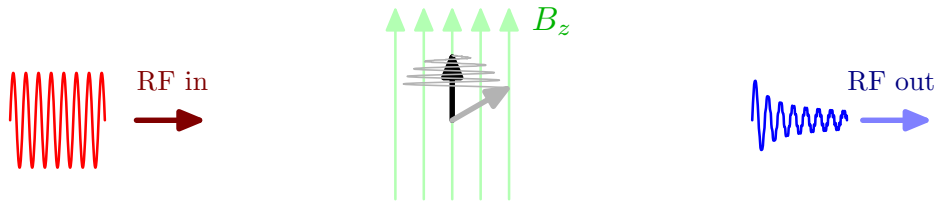


Figure 2.1: Schematic for the NMR principle (classic picture)

The concept consists of three steps: *Polarization*, *Excitation* and *Signal acquisition*:

1. *Polarization*

   Using an strong magnetic field $\vec{B}_z$, say in the direction $\vec{e}_z$, a major percentage of the nuclear magnetization is pointing in the direction $\vec{e}_z$. One call this process polarization.

2. *Excitation*

   One applies an electro-magnetic waveform in the xy-plane. If the waveform frequency equals the Larmor frequency, $\omega_{\text{in}} = \gamma \cdot |B_z|$, a resonance excitation will be induced, which means the magnetization vector precessing-wise flips away from the energy lowest state $|B_z| \cdot \vec{e}_z$.

   After this xy-plane electro-magnetic waveform is switched off, the nuclear magnetization has a non-zero angle to the external field. This is an excited state.

3. *Signal acquisition*

   The non-zero angle to the external magnetic field B_z results the precession motion around B_z at the Larmor frequency. This precession motion will induce an electro-magnetic waveform with the according Larmor frequency $\omega_{\text{out}} = \gamma \cdot |B_z|$.

   By detecting the frequency of the emitted (or absorbed) electro-magnetic waveform one can get the information about the γ value, (or the energy gap between the nuclear spin states $|\uparrow\rangle$ and $|\downarrow\rangle$). The γ value is tightly correlated with the microscopic structure of the material.

   By detecting the magnitude of the emitted waveform, one can get the information about the spin-density of the object. The temporal decay of the magnitude tells the relaxation properties of the spin system, which can be used to derive further microscopic structures.

   Remark: only the excited spins can be "seen".

The most medical imaging relevant Larmor frequencies are in the radio frequency band. So we call the electro-magnetic waveform in the xy direction the RF waveform.

2.3 From NMR to MRI, *how to enable resolution*

In the previous text we introduced the basic principle of nuclear magnetic resonance, in which we treated the object as a whole and trying to read its microscopic structure by observing its nuclear resonance excitation and relaxation process. However, this process doesn't offer any spatial resolution of the object.

For a successful application in medical imaging with the goal of a topological imaging of the inner structure of the human body, one need additional

mechanism to differentiate the tissue voxel from tissue voxel on different spatial locations within the object area. This means to deliver spatial resolution. For both excitation and signal acquisition steps, MRI requires enabling spatial resolution.

For the *excitation* process, the input electro-magnetic waveform has the same effect to all the local voxels. One controls them all together in the same way. In MRI one wants to find a way to fully control excitations of the local voxels individually. One call it the *spatial selective excitation*, which is the focus of this thesis.

For the *signal acquisition* process, the emitted waveform signal one detected is a superposition of the waveforms induced by different voxels in different positions. One needs to find a way to decode the information of the local voxels from the 'global' superposed waveform.

The solution for enabling the resolution for both excitation and signal acquisition stages is to add a linear varying component in the B_z. The basic idea is

- Both the resonance frequency for excitation and the magnetization precession frequency equal the Larmor frequency $\omega_L = \gamma \cdot B_z$.

- If one has a spatially varying $B_z = B_z(\xi)$, the spins in different spatial position will have different motion due to the different precession frequency. So one may have the possibility to control and read the motion of the voxels on different spatial positions individually.

- The most straightforward spatial varying type is 'linear'. One introduce thus the spatial gradient of the magnetic field $B_z(\xi)$.

$$B_z(\xi) = \langle \xi, g \rangle$$

For 1D it's sufficient to have a temporally constant slope. We call it the gradient g. However, for 2/3D imaging, one must go to temporally varying gradient $g = g(t)$ to enable a sufficient differentiation of the voxels distributed on 2/3D spatial region. We call this temporally varying control function $g(t)$ the gradient waveform.

We will introduce in more detail how to enabling resolution using $g(t)$ in §-3.1 for signal-acquisition and in §-4.1 for excitation.

The basic principle of MRI can be schematically represented in the (fig-2.2):

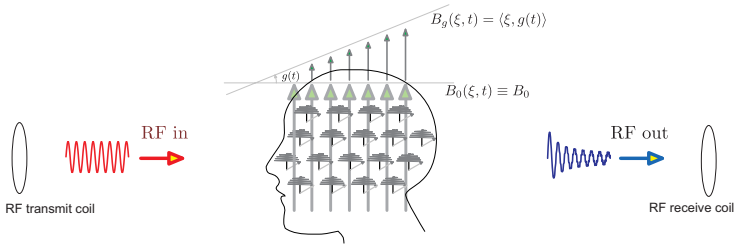


Figure 2.2: Schematic for the MRI principle

2.4 Two Main Tasks: *Excitation* & *Imaging*

Assume a sufficient polarization as given. Analog to the NMR principle, the MRI task can be roughly divided into two stages:

Excitation + Imaging .

Excitation We at first describe the *excitation stage*.

One has a nuclear spin system over certain spatial region $\mathbb{O} \subset \mathbb{C}$. We call it our object. The spatial coordinates for the individual pixels within this spin system is notated as $\xi \in \mathbb{O}$.

The magnetization vectors on each pixel obey the Bloch equation. It has a polarized initial state: $\forall \xi \in \mathbb{O}, \vec{M}(\xi, t_0) = (0, 0, 1)^T$

One has two control functions to influence the magnetization of the voxels: the electro magnetic RF waveform in xy-plane and the gradient waveform g as the spatial slope of B_z.

The task is to control the spatial magnetization profile of the nuclear spin system from the polarized initial state to a certain given spatial magnetization profile: $\vec{M}_{\text{target}} : \xi \mapsto \vec{M}_{\text{target}}(\xi) \in \mathbb{C}, \xi \in \mathbb{O}$. We call this profile the *target excitation profile*.

Imaging After the excitation stage, one want to 'read-out' the xy-magnetization M_{xy} as a function of ξ. This is the task of the *imaging stage*.

One has the magnetization profile, $M_{\text{xy}} : \xi \mapsto M_{\text{xy}}(\xi) \in \mathbb{C}, \xi \in \mathbb{O}$, to be reconstructed. The precession motion in the external magnetic field B_z induces superposition-wisely a RF waveform S, $S(t) = \int_{\mathbb{O}} M_{\text{xy}}(\xi) d\xi$. This RF waveform will be detected by the receive coil. During this process, the gradient waveform g can be applied as a control. The

signal function S of time t can be seen as a functional of the gradient waveform g of time t and the magnetization profile M_{xy} of ξ.

The task is to design the $g : [0, t_f] \mapsto \mathbb{C}$ in such a way that one can reconstruct the magnetization image $M_{\text{xy}} : \mathbb{O} \mapsto \mathbb{C}$ from the detected RF waveform signal $S : [0, t_f] \mapsto \mathbb{C}$.

The value of M_{xy} tells the spin density distribution within the object. From the temporal decaying of the magnetization value one gets the mapping of the relaxation-time parameters, $T_1 : \mathbb{O} \mapsto T_1(\xi) \in \mathbb{R}^+$ and $T_2 : \mathbb{O} \mapsto T_2(\xi) \in \mathbb{R}^+$, which correlated strongly with the biological properties of the tissue at the location ξ. In this way a MRI image can be produced (fig-2.3). (Using different excitation techniques, one can indeed measure separately the T_1 and the T_2. Since the two parameter correlated with different tissue properties, the resulting images are slightly different. Depending on what kind of tissue properties is interested in, one chooses respectively T_1-weighted or T_2-weighted image.)

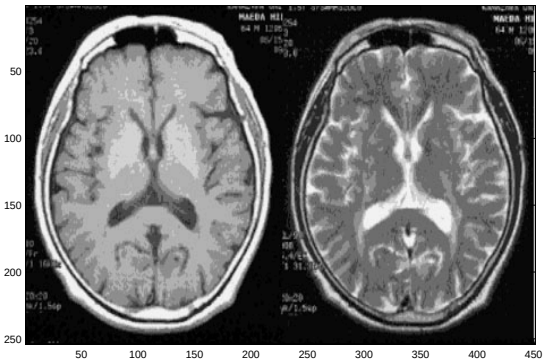


Figure 2.3: MRI brain image.
(*left:* T_1 weighted brain image; *right:* T_2 weighted brain image)

The relationship between the excitation and the imaging stages:
The xy-magnetization at certain spatial location is determined by two factors,

1. how strong is the local spin excitation,

2. the local tissue property, e.g. spin density, relaxation time, etc.

The factor-2, tissue property, is target information we want to reconstruct. The factor-1, excitation profile, can be seen as a possibility to decide, where

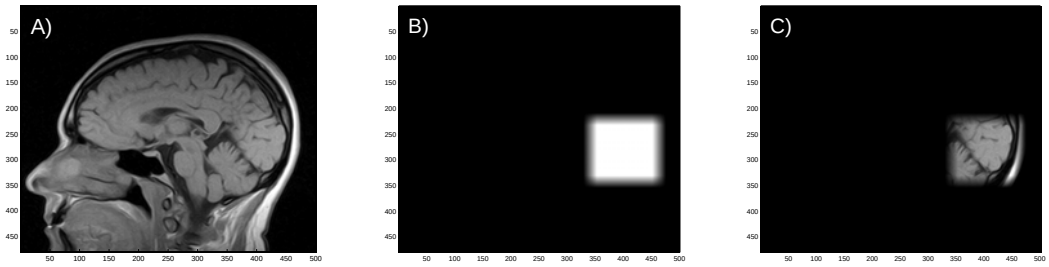


Figure 2.4: Excitation and Imaging
(*A):* the tissue topology, *B):* the excitation profile, *C):* the resulted image)

to look at. The MRI image can be seen as a superposition of the excitation profile and the tissue topology profile.

Figure-2.4 is an example to demonstrate the principle. Figure-2.4-A) shows a tissue topology of a human brain. Figure-2.4-B) is a excitation profile over the same region as A), where the bright region is the excited region. Since only the excited region can be seen, the resulting MRI image is a superposition of the tissue topology and the excitation profile, figure-2.4-C).

Physically there are many analogies between the excitation and the imaging process. One can be seen as an inverted process of the other. This symmetry between the two physical processes leads to strong analogy between the methods for the imaging stage and the excitation stage.

In the praxis there are strong wishes to shortening both the excitation and the imaging process. Although the motivations are slightly different(§-4.1), the methods for the acceleration in both stages turn to have many common part. In this thesis we concentrate on the acceleration of the excitation process.

Table-2.1 shows an overview of some related methods in both imaging and excitation sides. We will discuss them one by one in this chapter.

We discuss the standard method for imaging task in section 3.1. In section 3.2 and section 3.3 we will discuss the two important methods for accelerating the imaging process.

In section 4.1 we will discuss the standard approach for the spatial selective excitation problem. In section 4.2 we discuss the parallel transmit method in the aspect of accelerating the excitation process. It can be seen as an equivalence of the Parallel Receive method for the fast imaging.

After having the overview of the current methods in both imaging and excitation sides, we develop our own adaptive acceleration concept in section 4.3. Our concept can be seen as an analogue to the compressed sensing

	Imaging	Excitation
Standard Method	FT reconstruction	STA mehod
Acceleration via parallelization	Parallel RF Recieve	Parallel RF Transmit
Acceleration via prior-knowledge	Compressed Sensing	**Adaptive Sparse Excitation**

Table 2.1: The overview of the concepts

method for the fast imaging. Both enable the acceleration by utilizing the prior-knowledge. The information sparsity of the medical images and the excitation target profiles is essential for these both concepts.

Chapter 3

Imaging Problem

3.1 The Imaging Process

We devote this section to introduce the Fourier Transform reconstruction for imaging process to readout the individual voxel magnetization [LL99][BKZ04].

Since the focus of this thesis is the MR excitation process, this section and the following two sections are rather aimed to 1) set the view of the correspondence structure between the excitation process and the imaging process, 2) give a more complete picture of MRI for the non-MRI reader.

After recall the problem of imaging process, we will derive the method of Fourier Transform image-reconstruction, which reduces the task of imaging problem to a matrix inversion problem plus a so-called fast covering κ problem, where κ is a set of obligatory locations to be covered under velocity and acceleration constraints. At last we discuss the motivation and the challenge of the acceleration of the imaging process.

3.1.1 Recall the problem

We formulate the problem of imaging process as following:

$$(\boldsymbol{P}_{\text{Imaging}}) \quad \begin{cases} \text{One has the control function } g : t \mapsto g(t) \in \mathbb{C}; \\ \text{The signal } S \text{ is a functional of } g \text{ and } M_{xy}; \\ \\ \textit{Find } g, \text{ to } \textit{reconstruct } M_{xy}(\xi). \end{cases} \tag{3.1}$$

The imaging process can be schematically demonstrated in the fig-3.1.

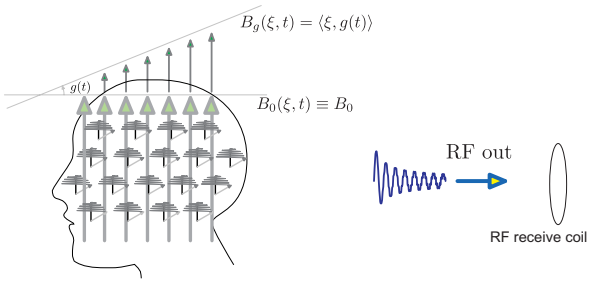


Figure 3.1: Schematic for the imaging process

3.1.2 The Fourier transform relationship

Let $\hat{M}_{xy} : \mathbb{O} \mapsto \mathbb{C}$ be the xy magnetization at $t_0 = 0$. By neglecting the relaxation effect, the motion of the $M_{xy}(\xi, t)$ is the precession motion caused by the external magnetic field $B_z(\xi, t) = B_0 + \langle g(t), \xi \rangle$. The according Larmor frequency is $\omega_0 = \gamma \cdot B_0 + \gamma \langle g(t), \xi \rangle$. The according magnetization motion is $M_{xy}(\xi, t) = \hat{M}_{xy} \cdot e^{i \cdot \gamma \cdot B_0 \cdot t} \cdot e^{i \cdot \gamma \cdot \int_0^t \langle g(\tau), \xi \rangle d\tau}$.

Since 1) the oscillation $e^{i \cdot \gamma \cdot B_0}$ is spatially and temporally constant and 2) $B_0 \gg \langle g(t), \xi \rangle$, it's convenient to make a notational simplification here to drop the term of $e^{i \cdot \gamma \cdot B_0}$ and keep in mind that $M_{xy}(\xi, t) = \hat{M}_{xy}(\xi) \cdot e^{i \cdot \gamma \cdot \int_0^t \langle g(\tau), \xi \rangle d\tau}$ is the envelope function of the true magnetization motion with a underlying carrier waveform $e^{i \cdot \gamma \cdot B_0}$.

The notational simplification can be physically understood as switching to the *rotating frame of reference* with rotating frequency of $\gamma \cdot B_0$. In MRI community one talks conventionally about the waveforms usually only in the *rotating frame of reference*. We will follow this convention in the rest of this work.

The detected RF waveform is the superposition of the $M_{xy}(\xi, t) \in \mathbb{C}$ over the whole object domain.

$$S(t) = \int_{\mathbb{O}} \hat{M}_{xy}(\xi) \cdot e^{i \cdot \gamma \cdot \int_0^t \langle g(\tau), \xi \rangle d\tau} d\xi$$

We see the clear similarity with the Fourier transform formulation proper definition of the frequency parameter:

$$k(t) := \gamma \int_0^t g(\tau) d\tau. \tag{3.2}$$

The quantity $k(t) \in \mathbb{C}$ is a very important concept in MRI. We call the according space the k space. We will have very similar concept also for the excitation problem.

Accordingly $M_{xy}(\xi, t) = \hat{M}_{xy}(\xi) \cdot e^{i\langle k(t),\xi\rangle}$, and

$$S(t) = \int_{\mathbb{O}} \hat{M}_{xy}(\xi) \cdot e^{i\cdot\langle k(t),\xi\rangle} d\xi$$

Since k is uniquely determined by t, one can express $S(t)$ as $S(k) = S(k(t))$. Then we get

$$S(k) = \int \hat{M}_{xy}(\xi) \cdot e^{i\cdot\langle k,\xi\rangle} d\xi \tag{3.3}$$

We identify the $M_{xy}(\xi)$ as the image to be reconstructed. With a finite image resolution, and a finite frequency resolution, one has a discrete Fourier Transform similar relationship, $S(k) = \sum_\xi e^{i\cdot\langle k,\xi\rangle} \cdot M_{xy}(\xi)$. We notate the pixel number in the image domain as $n := n_x \times n_y$ and the pixel number in the frequency domain as $m := m_x \times m_y$.

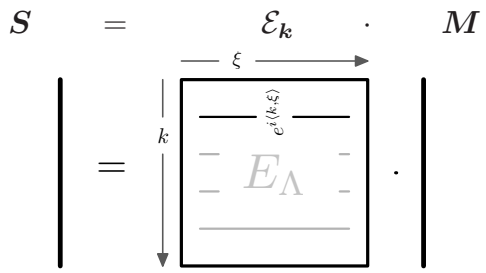
We define the so-called *encoding matrix*,

$$\boldsymbol{\mathcal{E}_k}(i,j) := e^{i\cdot\langle k_i,\xi_j\rangle} \in \mathbb{C}^{M\times N}, \quad \forall i : k_i \in \boldsymbol{k} \subset \mathbb{C}^M,$$

where $\boldsymbol{k} \in \mathbb{C}^M$ is the vector of the discrete k space sampling locations. Then (3.3) can be written as

$$\boldsymbol{S} = \boldsymbol{\mathcal{E}_k} \cdot \boldsymbol{M} \tag{3.4}$$

We call it the *encoding equation*. It can be seen as the core equation for the imaging process. The encoding equation can be schematically plotted as:



If the encoding matrix $\boldsymbol{\mathcal{E}_k}$ is invertible, one can reconstruct the image by

$$\boldsymbol{M} = \mathcal{E}_{\boldsymbol{k}}^{-1} \cdot \boldsymbol{S}$$

With the spatial grid ξ given, whether $\boldsymbol{\mathcal{E}_k}$ is invertible is decided by the k space sampling k.

From Shannon-Nyquist theorem we know if a set of discrete sampling locations fulfills the following two relationships

$$\begin{aligned} \delta\boldsymbol{\kappa} &= \tfrac{1}{\xi_{\max}} \\ \boldsymbol{\kappa}_{\max} &= \tfrac{1}{\delta\xi} \end{aligned} \quad ,$$

a successful reconstruction can be guaranteed. We'll call this special k space sampling set required by the Shannon Nyquist theorem the *Nyquist (sampling) grid*. With the Nyquist sampling grid the encoding equation is reduced to a discrete Fourier transform, $\mathcal{E}_{\boldsymbol{k}_{\mathrm{Nyq}}} = \mathrm{DFT}$. The reconstruction is reduced to an inverse discrete Fourier transform.

$$\begin{aligned} \boldsymbol{S} &= \mathrm{DFT} \cdot \boldsymbol{M} \\ \boldsymbol{M} &= \mathrm{DFT}^{-1} \cdot \boldsymbol{S} \end{aligned}$$

Numerically it can be performed economically by FFT.

3.1.3 Feasible k space traveling

We recall that the actual control function to be designed is the gradient waveform $g(t)$. The hardware realization of the gradient waveform gives the physical upper limit of the maximal gradient value and the maximal gradient slew rate:

$$\begin{aligned} g(t) &\leqslant G_{\max} \\ \dot{g}(t) &\leqslant S_{\max} \end{aligned}$$

In the equivalent k space picture, (3.2), it has the equivalent form

$$\dot{k}(t) \leqslant \gamma \cdot G_{\max} =: C_1 \tag{3.5}$$

$$\ddot{k}(t) \leqslant \gamma \cdot S_{\max} =: C_2 \tag{3.6}$$

In the previous subsection we derived a sufficient condition that with the Nyquist sampling grid being covered, one can always successfully reconstruct the desired image $\boldsymbol{M}$. Combined with the hardware feasibility constraints (3.5, 3.6), we have a problem of design a continuous fast traveling trajectory covering a given set of obligatory locations with the feasible constraints of traveling velocity and traveling acceleration. We call this problem the *fast covering* $\boldsymbol{\kappa}$ problem, where $\boldsymbol{\kappa}$ is the set of the obligatory locations to be covered. Here the obligatory location set takes the special form of Nyquist sampling grid: $\boldsymbol{\kappa}_{\mathrm{Nyq}}$. The problem will be introduced in (§-4.1)-(4.15) as $\boldsymbol{P}_{\mathrm{kTrav}}(\boldsymbol{\kappa}_{\boldsymbol{Nyq}})$.

At last, the feasible gradient waveform can be calculated straightforwardly by

$$g(t) = \dot{k}(t).$$

3.1.4 Accelerating the imaging process

One of the major challenges of MRI in medicine is how to shortening the total scanning duration. One also call this topic the fast imaging. The total scanning duration consists of the excitation duration plus the duration of the imaging stage. Usually the excitation duration only takes a fractional portion of the time. (The excitation acceleration is motivated by rather other reasons than the scanner duration shortening, more detail in Chapter-4.1). The scanning duration can be hence roughly seen as proportional to the duration of the imaging stage, which is the data acquisition time.

The data acquisition time is equal the duration of the k space trajectory covering the entire Nyquist sampling grid. Therefore a straightforward idea for acceleration is to undersample the Nyquist sampling grid, i.e. to reduce the number of the obligatory locations to be covered.

Generally, if one just simply undersamples the k space without any advanced treatment, one will get serious artifact by the reconstructed image. In the next two sections we will briefly discuss the two advanced approaches, the *parallel imaging* and the *compressed sensing*. Before that, we want to take a look at what issues and challenges we will have, if we just naively perform a direct undersampling of the Nyquist k grid.

Reducing the image pixel number such that the according Nyquist grid has less member is not an option, because

- If one reduces the spatial pixel number by increase the pixel size (spatial sampling distance), one just uncontrollable lost the detail image infomation of high spatial frequency part (the blurring artifact).

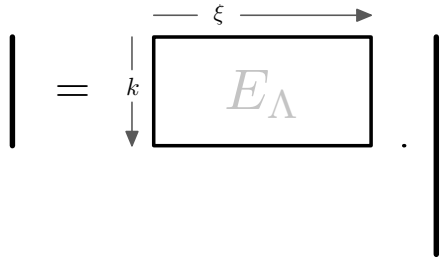
- If one reduces the FOV, the signal from the object part outside the FOV will be indistinguishably folded into the field of view (the aliasing artifact).

  This reduced FOV strategy could indeed have success, if one combines it together with the spatial selective excitation. (Section-5)

So we must face the situation of "#(k-sampling) < # pixels". The en-

coding matrix $\mathcal{E}_{\boldsymbol{k}}$ turns to have more columns than rows.

$$\boldsymbol{S} = \mathcal{E}_{\boldsymbol{k}} \cdot \boldsymbol{M} \tag{3.7}$$



(One can generally assume that all the row vectors are linear independent by an appropriate positioning of the sampling location.)

The task is to identify the magnetization vector $\boldsymbol{M} \in \mathbb{C}^N$ from the Domain of $\mathcal{E}_{\boldsymbol{k}}$ using the given $\boldsymbol{S} \in \mathbb{C}^M$ from the Range of $\mathcal{E}_{\boldsymbol{k}}$. The challenge we face here is that the Range is smaller than the Domain due to the undersampling of the Nyquist grid, $Range(\mathcal{E}_{\boldsymbol{k}}) \subset Dom(\mathcal{E}_{\boldsymbol{k}})$. This leads to the fact that alone the information from the signal vector from $Range(\mathcal{E}_{\boldsymbol{k}})$ is not sufficient to uniquely identify the the desired solution image vector $\boldsymbol{M} \in \mathbb{C}^N$.

So the key question is how to generate a encoding matrix $\mathcal{E}_{\boldsymbol{k}}$ with its Range has the same dimension as its Domain :

$$Range(\mathcal{E}_{\boldsymbol{k}}) \supseteq Dom(\mathcal{E}_{\boldsymbol{k}})$$

There are two alternative directions one can try:

1. artificially expand the $Range(\mathcal{E}_{\boldsymbol{k}})$ despite of the undersampled Nyquist grid;

2. bring in additional information to scale down the $Dom(\mathcal{E}_{\boldsymbol{k}})$.

The two established approaches of enabling Nyquist grid undersampling can be seen as exactly following the two directions above. The *parallel imaging* approach provides the parallelization to artificially expand the $Range(\mathcal{E}_{\boldsymbol{k}})$. The *compressed sensing* approach introduces additional prior-knowledge to scale down the $Dom(\mathcal{E}_{\boldsymbol{k}})$.

In the following two chapters we will introduce these two approaches in turn.

3.2 Parallel Imaging

In this section we will discuss the first one of the two major established approaches of enabling undersampling in the MR imaging process: the *Parallel Imaging*, also called parallel receive [PWSB99] [GJH+02] [Pru04].

We will at first introduce the additional physical setting and the intuitive motivation for this concept, then give a mathematical explanation of why it works. We will skip many important technical details and introduce only the very simplified core concept.

3.2.1 Physical setting and the intuitive motivation

At first we want to introduce *the receive sensitivity profile.* The schematic plot (fig-3.1) shows the normal imaging process. Here one has only a single RF receive coil. The RF waveform detected by the receive coil is a superposition of the signals from all the voxels over the whole object region.

$$S(t) = \int_{\mathbb{O}} M_{xy}(\xi, t) d\xi$$

In general, the voxels of different locations will be observed by the receive coil with different sensitivity. The superposition is indeed a weighted superposition.

$$S(t) = \int_{\mathbb{O}} s(\xi) \cdot M_{xy}(\xi, t) d\xi$$

We call this spatial weighting function the receive sensitivity profile $s : \mathbb{O} \mapsto \mathbb{C}$. The hardware geometry and the positioning of the receive coil are the essential factors to influence the sensitivity profile.

For the single receive coil setting like in fig-3.1, the receive coil is usually constructed in such a way that its sensitivity profile is approximately constant over the whole region (fig-3.3-left). In the parallel imaging, instead of one single RF receive detector one has a series of receive detectors. Each of them is intentionally constructed to have different receive sensitivity profile, e.g. fig-3.3-right. The setting can be schematically described in the plot (fig-3.2):

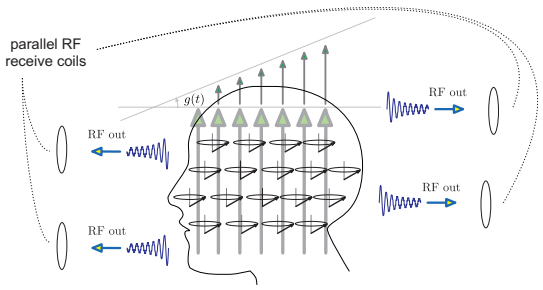


Figure 3.2: Schematic for the parallel imaging setting

An intuitive motivation of the parallel receive concept can be demonstrated in the following theoretical limit case. One has four ideal receive-detectors. Each of them can only detect the signals from a restricted region of the whole object (fig-3.3-right). The union of these four partial sensitive regions however covers the whole FOV.

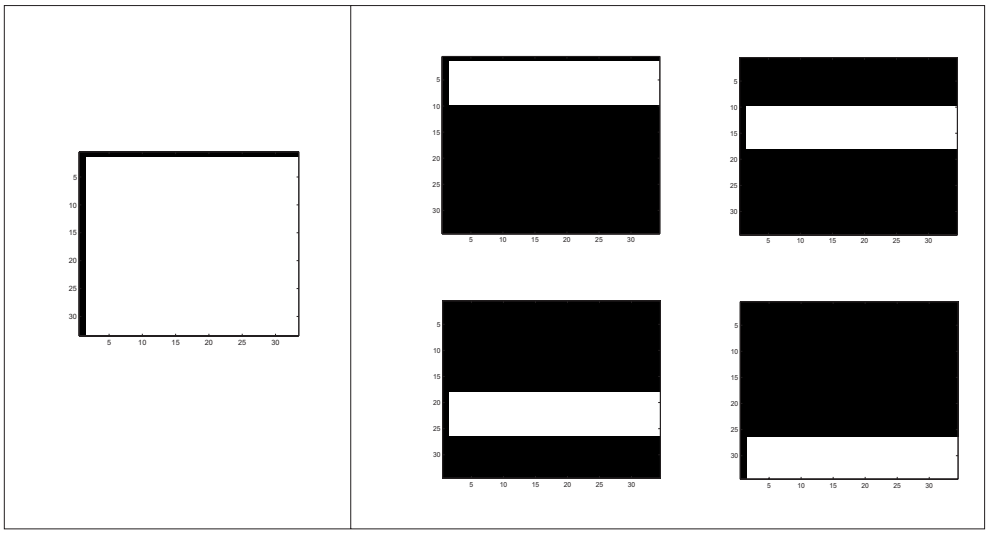


Figure 3.3: The ideal (parallel) receive spatial sensitivity profiles
(*Left:* homogeneous single receive channel sensitivity.
Right: four ideally localized receive channel sensitivity profiles.)

One can receive the signals in parallel using all the detectors. Since the all detectors have now smaller (but different) responsible regions, the according Nyquist sampling grid has hence larger sampling distance (fig-3.4-right) compared with the original one (fig-3.4-left). In this way one get the capability of undersampling the k space without quality reduction.

In the praxis however, one will never have such ideal receive sensitivity profiles as in (fig-3.3-right). But with the compromise on the undersampling rate, one can still achieve the undersampling with reasonable quality. For example 3× undersampling with 6 receive coils.

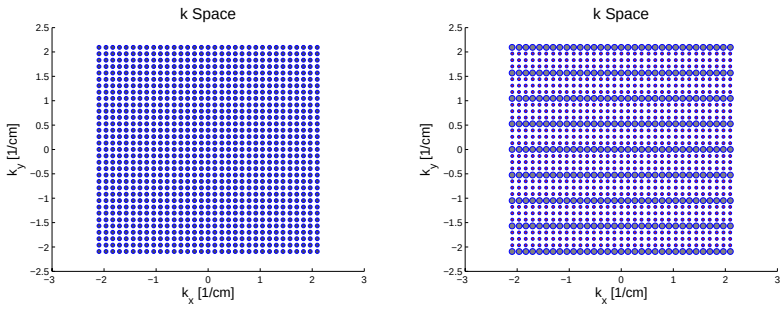
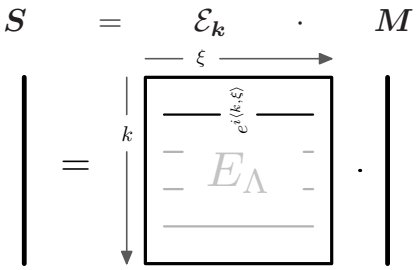


Figure 3.4: Undersampling the Nyquist due to parallel setting.
(*Left:* Nyquist fullsampling. *Right:* the small-dots are the Nyquist grid, the big-dots are the undersampled pattern.)

3.2.2 Mathematical justification

To concretely explain the principle of parallel imaging concept, we start with the encoding equation:



If $\mathcal{E}_{\boldsymbol{k}}$ is a regular invertible matrix, the magnetization profile can be reconstructed by $\boldsymbol{M} = \mathcal{E}_{\boldsymbol{k}}^{-1} \cdot \boldsymbol{S}$. The core criterion for if it can also be done by general case is decided by the size of the range and domain of the $\mathcal{E}_{\boldsymbol{k}}$. The range of $\mathcal{E}_{\boldsymbol{k}}$ is the signal space. The domain of $\mathcal{E}_{\boldsymbol{k}}$ is the space of the images (or the magnetization profiles) to be reconstructed. We call it the magnetization space. The condition of if one can correctly reconstruct the $\boldsymbol{M}$ from $\boldsymbol{S}$ is

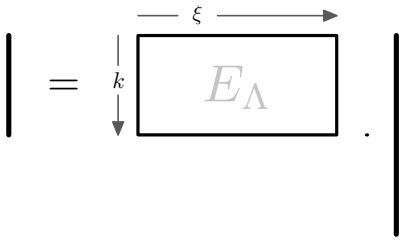
if the range of $\mathcal{E}_{\boldsymbol{k}}$ (the signal space) can cover the domain of $\mathcal{E}_{\boldsymbol{k}}$ (the magnetization space)

So, an equal- or over-determined encoding equation is wished, an under-determined encoding equation is problematic.

With an undersampled Nyquist sampling grid, $\#(\boldsymbol{k}) < \#(\boldsymbol{k}_{\text{Nyq}}) = N$, one will get a rectangular encoding matrix $\mathcal{E}_{\boldsymbol{k}} \in \mathbb{C}^{L \times N}$, $L < N$. The encoding

matrix is basically underdetermined.

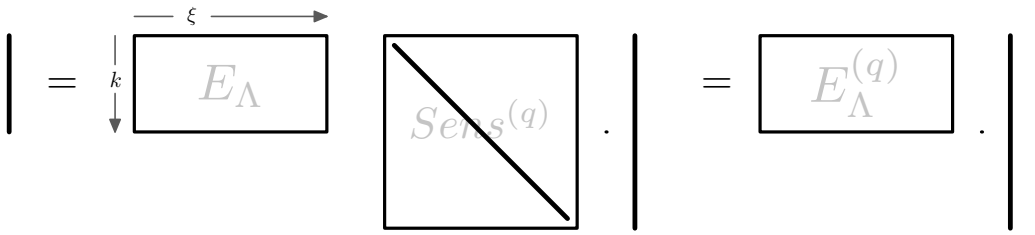
$$\boldsymbol{S} = \mathcal{E}_{\boldsymbol{k}} \cdot \boldsymbol{M} \tag{3.8}$$



In this case the inversion without any advanced treatment is a ill-posed problem due to the underdetermined nature. The way of parallel imaging to release this problem is to "artificially" increase the column space of $\mathcal{E}_{\boldsymbol{k}}$ (the signal space).

Say, there are C parallel receive coils, whose sensitivity profiles are sufficiently different from each other. We also call them C parallel receive channels. One will get totally C different signal vectors from the different receive channels, $\boldsymbol{S}^{(c)}$, $c = 1 : C$. Each channel has its own individual encoding equation:

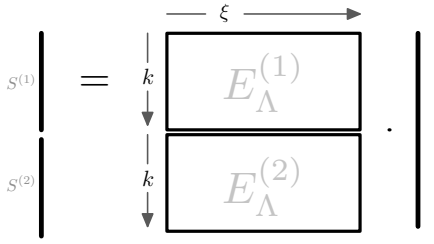
$$\boldsymbol{S}^{(c)} = \mathcal{E}_{\boldsymbol{k}} \cdot \mathcal{S}^{(c)} \cdot \boldsymbol{M} = \hat{\mathcal{E}}^{(c)}_{\boldsymbol{k}} \cdot \boldsymbol{M}$$



Due to the parallel receive setting, one has up to C-fold redundant information. The idea is to use this information redundancy to compensate the information deficit caused by the undersampling.

One can write the new composite encoding equation, e.g. for a case of $C = 2$:

$$\boldsymbol{S} = \hat{\mathcal{E}}_{\boldsymbol{k}} \cdot \boldsymbol{M}$$

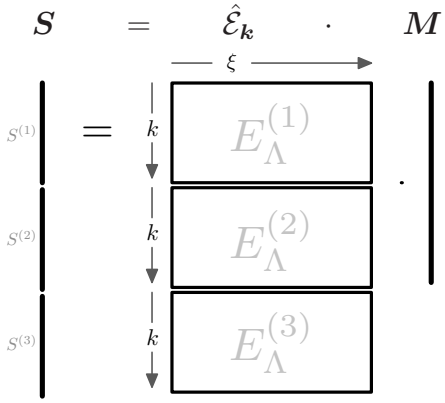


Where: $\boldsymbol{S}$ is the in two direction concatenated signal vector over all receive channels. $\hat{\mathcal{E}}_{\boldsymbol{k}}$ is the in row direction concatenated encoding matrix over all receive channels.

Whether the new defined $\hat{\mathcal{E}}_{\boldsymbol{k}}$ is invertible depends on 1) the undersampling, 2) the number of independent receive channels and their sensitivity profiles.

In the theoretical limit case as in (fig-3.3-right), the composite encoding matrix will remain unitary. Thus it's perfectly invertible with unit condition number. Therefore it allows a undersampling rate as high as the number of the parallel receive channels.

In the praxis, the realistic sensitivity profiles is far from the ideal one as in (fig-3.3-right). It reduce the row orthogonality of the encoding matrix. Some of the row vectors of $\hat{\mathcal{E}}_{\boldsymbol{k}}$ are linear dependent. Thus the dimension of the row space of $\hat{\mathcal{E}}_{\boldsymbol{k}}$ will be smaller than $C \cdot \#(\boldsymbol{k})$. Therefore, in the praxis the allowed undersampling rate is usually smaller than the number of the independent receive channels, e.g. 2×undersampling using 3 receive channels.



An important issue here to address is the condition number of the encoding matrix. For realistic sensitivity profiles, the condition number of the concatenated encoding matrix $\hat{\mathcal{E}}_{\boldsymbol{k}}$ is usually quite high. The bad conditionedness of the encoding matrix will cause a unwished effect that the noise in the signal $\boldsymbol{S}$ got amplified through the encoding process. So one has the problem of having a reduced Signal to Noise Ratio (SNR) due to the increased condition number of $\hat{\mathcal{E}}_{\boldsymbol{k}}$. It can be seen as a cost of enabling undersampling. In the praxis, to keep the condition number under control, one has to reduce the undersampling rate or, equivalently, increase the number of independent receives channels.

In MR parallel imaging terminology this SNR amplification effect due to the bad conditioned encoding matrix is notated as the g-factor.

For conclusion, the core principle of *parallel imaging* is:

Utilizing the information redundancy from parallelization to repair the information deficit due to the undersampling. Key condition for that is the numerical invertibility of the resulting composite encoding matrix.

3.3 Compressed Sensing

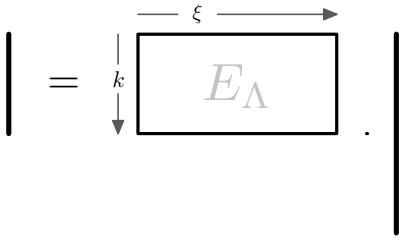
In this section we will discuss the second one of the two major established approaches of enabling undersampling in the MR imaging process: *compressed sensing for MRI imaging*, [LDP07], [LDSP08], [GBK08], [JSN+09], [LKS+09]. It is based on the rapidly developed topic of *Compressed Sensing* in applied mathematics, [CRT06], [Don06].

Like the section for parallel receive, the motivation of this section is not to discuss the technique of compressed sensing and its application in MRI imaging stage itself, but to high-light its basic idea and prepare the correspondent relationship with the excitation side. Thus many important technical details are skipped, only the very simplified basic ideas are introduced. Some further technical detail will also be discussed in the next chapter for sparse approximation.

3.3.1 Compressed Sensing(CS) in MRI

At first we recall the issue by the naive directly undersampling of the Nyquist k grid.

$$\boldsymbol{S} \quad = \quad \hat{\mathcal{E}}_{\boldsymbol{k}} \quad \cdot \quad \boldsymbol{M} \tag{3.9}$$



The major challenge here is that the magnetization space, $\mathrm{Dom}(\hat{\mathcal{E}}_{\boldsymbol{k}})$, cannot be completely covered by the signal space, $\mathfrak{S}$. The signal space is equal the range of the encoding matrix, $\mathrm{Range}(\hat{\mathcal{E}}_{\boldsymbol{k}})$. It means that the solution-space of (3.9) can only be reduced from $\mathrm{Dom}(\hat{\mathcal{E}}_{\boldsymbol{k}}) = \mathbb{C}^N$ to a finite dimensional subspace of $\mathbb{C}^N$, but not to a unique point in $\mathrm{Dom}(\hat{\mathcal{E}}_{\boldsymbol{k}})$. So one has infinite many number of "correct" solutions subject (3.9).

The parallel imaging concept follows the strategy to increase the signal space, $\mathfrak{S} = \mathrm{Range}(\hat{\mathcal{E}}_{\boldsymbol{k}})$, such one can have enough information to reduce the solution space to a unique point in $\mathrm{Dom}(\hat{\mathcal{E}}_{\boldsymbol{k}})$.

The alternative strategy to cover the magnetization space by the signal space,

$$\mathfrak{S} \supseteq \mathrm{Dom}(\hat{\mathcal{E}}_{\boldsymbol{k}})$$

would be *to reduce the* $\mathrm{Dom}(\hat{\mathcal{E}}_{\boldsymbol{k}})$. Or differently formulated, *one could bring-in additional criteria enabling selecting out the desired* true *solution from the infinitive many possible solutions*. In this way one could still successfully reconstruct the image despite of the undersampling.

The realization of this alternative strategy can be reduces to two questions:

1. What's the additional criteria, allowing to pick up the true unique image from the infinite group of possible solutions. This is equivalent to ask *what makes our image special?*

2. How to technically realize the selection mechanism regarding to the additional criterion.

The answer for the first question is *the information low dimensionality*, which distinguishes the true image from the other solution candidates. Empirically, one can say most natural images have very low information amount compared with the number of pixels presenting them. In certain transformed domain it can be represented with very low number of basis vectors. One calls this property *transformed sparsity*. Empirically, this transformed sparsity criterion is sufficient to uniquely identify the true image to be reconstructed.

The second question, how to technically realize the strategy of utilizing the sparsity prior knowledge to enable the undersampling, leads us to the topic of compressed sensing. [CRT06], [Don06]

Compressed sensing consists of two steps of tasks. Firstly, how to design the undersampling pattern to maintain the essential information of the source image despite of the undersampling. Secondly, how to reconstruct the image from 1)the sampled data and 2)the additional transformed sparsity criterion.

The answer for the first task of sampling strategy is the *random sampling*. If the sampling pattern has sufficient randomness, then the image is reconstructable despite of the according undersampling rate.

In MRI the pure random sampling is not efficient because of the hardware constraints of the k space traveling. Thus, instead of a pure random pattern, one uses a randomly disturbed regular trajectory. This is a compromise between the optimal information collection and the k space traveling efficiency.

The second task of sparsity constraint reconstruction leads to the so-called *overcomplete sparse approximation* problem. The sparsity of a vector

is usually notated with the so-called l_0 semi-norm. The l_0-norm, $\|\boldsymbol{v}\|_0$, is defined as the number of the non-zero entries of the input vector $\boldsymbol{v}$. The task of the sparsity constraint reconstruction can then be formulated as

$$(\boldsymbol{P}_{\text{CS-MRI-}L_0}) \quad \begin{cases} \text{minimize } \|M\|_0 \\ s.t. \quad \boldsymbol{S} = \hat{\mathcal{E}}_{\boldsymbol{k}} \cdot \boldsymbol{M} \end{cases} \tag{3.10}$$

This problem is of combinatorial nature and in general NP hard.

A very popular approximate approach for that in the compressed sensing community is to approximate the original combinatorial l_0-regularized underdetermined least squares problem with the l_1-regularized underdetermined least squares problem.

$$(\boldsymbol{P}_{\text{CS-MRI-}L_1}) \quad \begin{cases} \text{minimize } \|M\|_1 \\ s.t. \quad \boldsymbol{S} = \hat{\mathcal{E}}_{\boldsymbol{k}} \cdot \boldsymbol{M} \end{cases} \tag{3.11}$$

It can be reformulated as a linear programming problem, which is a convex optimization problem. One can solve it with standard software. The l_1-regularization approach is also referred to as BP method (Basis Pursuit).

We will discuss the overcomplete sparse approximation problem in more detail in next chapter.

The compressed sensing approach has demonstrated great success in MR imaging process. As an independent mechanism, one can achieve significant acceleration without a parallel imaging setting. The reader can find more details about compressed sensing in the work of Candes et.al. [CRT06], and Donoho, [Don06]. For applying the compressed sensing in MRI the works of Lustig provide detailed information, [LDP07], [LDSP08].

3.3.2 Combining Parallel Imaging with CS

The parallel imaging method and the compressed sensing method represent two independent mechanisms to enable the undersampling of the Nyquist grid in k space. It's desirable to combine the both concepts together to achieve a maximal undersampling capability.

Indeed it turns to be a highly researched topic in the MRI community in the recent two years, [Kin08], [MHB08], [LKS+09]. The underlying technique comes from the recent development in the compressed sensing community, the so-called simultaneous sparse approximation, [TGS06], [Tro06]. By simultaneous sparse approximation one wants to simultaneously, sparsely linear represent a series of target signals using the same overcomplete collection of the elementary vectors.

The recent work of [Kin08], [MHB08], [LKS+09] demonstrated successful integration of parallel imaging and the compressed sensing. Indeed one can achieve significant further acceleration of the imaging process compared with applying only the one of two mechanisms separately.

Chapter 4

Excitation Problem

4.1 Spatial Selective Excitation Problem (SSEP)

In the previous three sections we've discussed the tasks and the methods of the imaging stage. In the next two sections we are going to discuss the tasks and the methods in the excitation stage, the spatial selective excitation problem.

To simplify the notation, we sometimes abbreviate the "Spatial Selective Excitation Problem" as SSE-Problem or SSEP.

The goal of this section is to derive the spatial selective excitation problem to a linear least squares formulation, [PNM89a].

In subsection-1 we will discuss the motivation of the Spatial Selective Excitation Problem and the formally formulate the problem, $\boldsymbol{P}_{\text{SSEP}}$.

In subsection-2 we will derive the conventional method based on the so-called *STA-model*, which reduce the $\boldsymbol{P}_{\text{SSEP}}$ to fast covering $\boldsymbol{\kappa}_{Nyq}$ problem, $\boldsymbol{P}_{\text{kTrav}}(\boldsymbol{\kappa}_{Nyq})$, plus a linear least squares problem to design the RF waveform, $\boldsymbol{P}_{\text{RF-Design}}$.

In subsection-3 we will give two examples to demonstrate the STA approach.

In subsection-4 we will discuss the motivation and the challenges of accelerating the spatial selective excitation process.

4.1.1 The spatial selective excitation problem

The Motivation

Depending on the concrete application situation, one wants to excite the nuclear spins on different spatial locations with different target excitation

magnitude. We call the given spatially varying excitation magnitudes the *target excitation profile*. The control function to achieve that is the RF waveform and the gradient waveform. The task is to design the RF and g to reach the target excitation profile.

Schematically it can be demonstrated in fig-4.1.

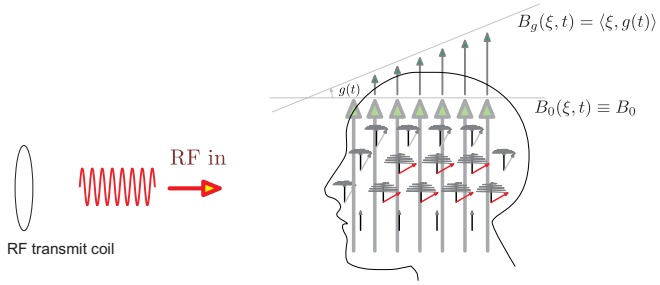


Figure 4.1: Schematic for the spatial selective excitation process

The general spatial selective excitation problem turns to gain more and more interest in the recent years. The motivations for the general spatial selective excitation problem are basically of the following two points:

1. localized excitation:

   By only exciting a well known small region of interest out of the whole object region and *not exciting* elsewhere, one can reduce the effort of the imaging stage. If staying with the same resolution, one can now reduce the imaging acquisition time, because the total pixel number to encoding is now smaller. Or, if one is willing to "pay" the same amount of imaging encoding steps, one can achieve an accordingly higher resolution.

   This principle can be demonstrated in the following figures.

   By "not exciting" certain uninterested region, which has artifact source, (e.g. a beating heart causing motion artifact), one would have the possibility to reduce the image artifact.

2. homogenization of the excitation

   In the praxis the transmit rf coils has no spatially homogeneous response property. Especially in the high B_0 field setting, the spatial inhomogeneity of the RF waveform is so high that for some region there is no enough signal intensity to reconstruct the image.

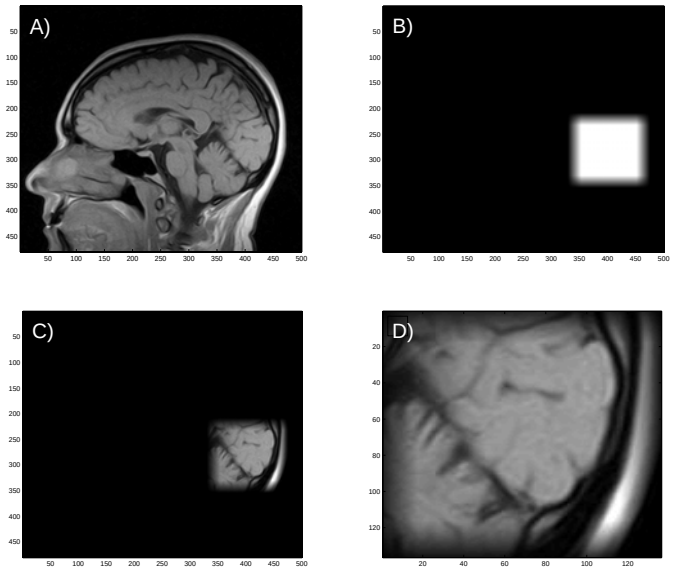


Figure 4.2: Principle of small region excitation.
A): the tissue topology, *B):* the excitation profile, *C):* the resulted image, *D):* the "zoom-in" FOV

One approach to release this issue is to selectively excite the object with the inversion of the RF spatial sensitivity as the target excitation profile. In this way the low sensitive region will be intentionally excited with larger magnitude to compensate the low sensitivity effect.

The Problem

At first we make an assumption that the duration of the excitation process is short enough that the relaxation effect can be sufficiently ignored. This means we can neglect the last two terms for relaxation in the Bloch equation.

So the Bloch equation (2.1) can be simplified to

$$\dot{\vec{M}}(\xi,t) = \vec{M}(\xi,t) \times \gamma \cdot \vec{B}(\xi,t)$$

where $\vec{B}(\xi,t)$ is the 3D magnetization vector, $B(\xi,t) = \Big(\mathrm{RF}_x(t), \quad \mathrm{RF}_y(t), \quad \langle g(t), \xi \rangle \Big)^T$.

Write the vector cross-product '$\times$' in matrix form, we get:

$$\begin{pmatrix} \dot{M}_x(\xi,t) \\ \dot{M}_y(\xi,t) \\ \dot{M}_z(\xi,t) \end{pmatrix} = \gamma \cdot \begin{pmatrix} 0 & B_z(\xi,t) & -B_y(\xi,t) \\ -B_z(\xi,t) & 0 & B_x(\xi,t) \\ B_y(\xi,t) & -B_x(\xi,t) & 0 \end{pmatrix} \begin{pmatrix} M_x(t) \\ M_y(t) \\ M_z(t) \end{pmatrix}, \tag{4.1}$$

Where:

$$\begin{aligned} B_x(\xi,t) &:= \text{real}(s(\xi) \cdot \text{RF}(t)) \\ B_y(\xi,t) &:= \text{imag}(s(\xi) \cdot \text{RF}(t)) \\ B_z(\xi,t) &:= \langle \vec{g}(t), \vec{\xi} \rangle \end{aligned}$$

Initial value is the polarized state:

$$\forall \xi, \vec{M}(\xi, t_0) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

With the assumption of neglecting the relaxation process and the notation used above, we give the formal problem statement of the **Spatial Selective Excitation Problem**.

$$(\boldsymbol{P}_{\text{SSEP}}) \begin{cases} \text{Given:} & \begin{cases} \boldsymbol{p} \in \mathbb{C}^N, \text{(target profile)} \\ \boldsymbol{s} \in \mathbb{C}^N, \text{(sensitivity profile)} \\ \text{TOL (user-tolerance)} \end{cases} \\ \text{Find:} & RF : [0, t_f] \mapsto \mathbb{C},\ g : [0, t_f] \mapsto \mathbb{C} \\ \text{Subject to:} & \begin{cases} \boldsymbol{M} : [0, t_f] \mapsto \mathbb{C}^N \text{satisfies the Bloch eq.(4.1)}, \\ \begin{cases} \|\dot{k}\|_\infty \leqslant C_1 \\ \|\ddot{k}\|_\infty \leqslant C_2 \end{cases} \text{(feasibility constraints).} \\ \exists t_f : \|\boldsymbol{M}(t_f) - \boldsymbol{p}\|_2 \leqslant \text{TOL} \end{cases} \end{cases} \tag{4.2}$$

(4.3)

The problem $\boldsymbol{P}_{\text{SSEP}}$ can be seen as a control problem. Its controllability has been proven in the work of Conolly et.al., [CNM86].

4.1.2 The conventional approach (Small tip angle regime, single RF channel)

In previous subsection we've formally formulated the spatial selective excitation problem. In this section we introduce the conventional approach to this problem.

We call this method the Small-Tip-Angle approach (STA) due to its core assumption of the small tip angle of the magnetization vector. This method is introduced by J. Pauly at 1989, [PNM89a]. Since then it is the most established design method for multi-dimensional selective excitation design problem.

Let $M_{xy}(\xi,t) := M_x(\xi,t)+i\cdot M_y(\xi,t)$ be the xy magnetization in complex form, and let $B_{xy}(\xi,t) := B_x(\xi,t)+i\cdot B_y(\xi,t) = s(\xi)\cdot \mathrm{RF}(t)$ be the xy magnetic field. The relaxation-free Bloch equation (4.1) turns to be:

$$\begin{aligned} \dot{M}_{xy}(\xi,t) &= -i\langle g(t),\xi\rangle B(\xi,t) + iB(\xi,t)M_z(\xi,t) && (4.4)\\ \dot{M}_z(\xi,t) &= -iB_{xy}(\xi,t)M_{xy}(\xi,t) && (4.5) \end{aligned}$$

due to the norm conservation property of the non-relaxation Bloch equation:

$$\begin{aligned} M_z &= \sqrt{1-|M_{xy}|^2} \\ &= 1+\mathcal{O}(|M_{xy}|^2) \end{aligned}$$

In the first order approximation of $|M_{xy}| \ll 1$:

$$\begin{aligned} M_z &\approx 1 \\ \dot{M}_z &\approx 0 \end{aligned}$$

With this approximation the equation (4.4) and (4.5) can be successfully decoupled. This approximation is physically an assumption that the flip angle between the magnetization vector $\vec{M}$ and the ground state $\vec{e}_z$ is small:

$$\|\vec{M}\|_2 \equiv 1: \quad \theta = \sin(\theta) = \tfrac{M_{xy}}{\|\vec{M}\|_2} = M_{xy} \quad .$$

We call it the STA assumption (Small-Tip-Angle-assumption).

The relaxation-free Bloch equation then turns to be:

$$\dot{M}_{xy}(\xi,t) = -i\cdot\langle g(t),\xi\rangle\cdot M_{xy}(\xi,t) + i\cdot B_{xy}(\xi,t) \qquad (4.6)$$

This is a linear ordinary differential equation with initial value: $M_{xy}(\xi, t_0) = 0, \forall \xi \in \mathbb{O}$. Solving (4.6) one gets:

$$M_{xy}(\xi, t_f) = i \int_0^{t_f} B_{xy}(\xi, t) \cdot e^{-i\xi^T \int_t^{t_f} g(\tau) d\tau} dt \tag{4.7}$$

Eq-(4.7) has strong similarity with Fourier Transform. One defines the frequency parameter $k \in \mathbb{C}$:

$$k(t) := \int_{t_f}^{t} g(\tau) d\tau \tag{4.8}$$

then eq-(4.7) becomes

$$M_{xy}(\xi, t_f) = i \cdot \int_0^{t_f} B_{xy}(\xi, t) \cdot e^{-i\langle \xi^T, k \rangle} dt \tag{4.9}$$

With a discretization in both spatial and frequency domain, (4.9) turns to be

$$\boldsymbol{M} = \mathcal{S} \cdot \mathcal{T} \cdot \boldsymbol{b} \tag{4.10}$$

where $\boldsymbol{M} \in \mathbb{C}^N$ is the discrete spatial xy magnetization profile, $\boldsymbol{b} \in \mathbb{C}^M$ is the discrete RF waveform as a function of the frequency domain sampling, $\mathcal{S} := \mathrm{Diag}(\boldsymbol{s})$, where $\boldsymbol{s} \in \mathbb{C}^N$ is the discrete form of $s(\xi)$, and $\mathcal{T} \in \mathbb{C}^{n \times m}$ is the so-called transmit matrix, which has a DFT similar form:

$$\mathcal{T} := \begin{pmatrix} e^{i\xi_1 \boldsymbol{k}_1} & e^{i\xi_1 \boldsymbol{k}_2} & \dots & e^{i\xi_1 \boldsymbol{k}_m} \\ e^{i\xi_2 \boldsymbol{k}_1} & e^{i\xi_2 \boldsymbol{k}_2} & \dots & e^{i\xi_2 \boldsymbol{k}_m} \\ \vdots & \vdots & & \vdots \\ e^{i\xi_N \boldsymbol{k}_1} & e^{i\xi_N \boldsymbol{k}_2} & \dots & e^{i\xi_N \boldsymbol{k}_m} \end{pmatrix} \tag{4.11}$$

The frequency domain sampling k is a function of time t, $k : t \mapsto k(t) \in \mathbb{C}$. The function k is uniquely determined by g, and verse vise. Therefore the task of designing RF and g in $\boldsymbol{P}_{\mathbf{SSEP}}$ can then be transfered to the task of finding k and RF, in discrete form $\boldsymbol{k}$ and $\boldsymbol{b}$, which satisfies

$$\|\boldsymbol{p} - \mathcal{S} \cdot \mathcal{T} \cdot \boldsymbol{b}\|_2 < \mathrm{TOL}\,. \tag{4.12}$$

We also write it as

$$\boldsymbol{p} \overset{\mathrm{TOL}}{=} \hat{\mathcal{T}} \cdot \boldsymbol{b}, \tag{4.13}$$

where $\hat{\mathcal{T}} = \mathcal{S} \cdot \mathcal{T}$.

due to the similarity with the Fourier Transform, the Shannon-Nyquist theorem can help us to determine a sufficient set of k space sampling locations, which guarantees a solution of (4.12).

Shannon-Nyquist theorem tells us that if one choose the k locations fulfill the following two criteria:

$$\text{(Shannon-Nyquist-criteria:)} \quad \begin{cases} \delta k &= 2\pi/\,\mathrm{FOV} \\ (k_{\max} - k_{\min}) &= 2\pi/\delta\xi \end{cases}, \tag{4.14}$$

then the existence of a solution of (4.12) is guaranteed.

In the rest of this work we call (4.14) the Shannon-Nyquist criteria. We call the set of the k locations fulfilling the Shannon-Nyquist criteria the Nyquist-grid. We notate the k space sampling locations followed by the Shannon Nyquist criterion as $\boldsymbol{\kappa}_{\mathrm{nyq}}$.

After getting the Shannon Nyquist criterion as the sufficient condition for the k space sampling, we have the following two steps to complete the task of $\boldsymbol{P}_{\mathbf{SSEP}}$:

I. Subject to the physical constraints (3.5) and (3.6), one wants to design a feassible $k(t)$ — thus also the gradient waveform $g(t)$ — from $\boldsymbol{\kappa}_{\mathrm{nyq}}$. The $g(t)$ and thus $k(t)$ are indeed then discretely realized in hardware level. We notate their discrete form accordingly as $\boldsymbol{g} \in \mathbb{C}^M$ and $\boldsymbol{k} \in \mathbb{C}^M$.

II. With the resulted k space sampling trajectory, the next task is to determine the correct RF waveform vector $\boldsymbol{b}$ according (4.12). It's a linear least squares problem.

One can formally formulate tasks of the step-I and -II accordingly as $\boldsymbol{P}_{\mathrm{kTrav}}(\boldsymbol{\kappa}_{Nyq})$ and $\boldsymbol{P}_{\text{RF-Design}}$.

The fast covering $\boldsymbol{\kappa}$ problem, $\boldsymbol{P}_{\mathrm{kTrav}}$, has the task of finding a continuous fast traveling trajectory covering a given set of obligatory locations with the feasible constraints of traveling velocity and traveling acceleration:

$$(\boldsymbol{P}_{\mathrm{kTrav}}) \quad \begin{cases} \textit{find:} \quad k : [0, t_f] \mapsto \mathbb{C} \\ \quad \textit{s.t.} \quad \begin{cases} t_f = \mathit{minimal} & \text{(minimal time)} \\ \mathit{range}(k) \supset \boldsymbol{\kappa} & \text{(covering)} \\ \\ \|\dot{k}\|_\infty \leqslant C_1 \\ \|\ddot{k}\|_\infty \leqslant C_2 \end{cases} \text{(feasibility constraints)} \; . \end{cases} \tag{4.15}$$

With the notation $\boldsymbol{P}_{\text{kTrav}}(\boldsymbol{\kappa}_{Nyq})$ we mean the special case of $\boldsymbol{P}_{\text{kTrav}}$ with the very regular Nyquist grid as the input obligatory locations.

Completely solving the $\boldsymbol{P}_{\text{kTrav}}$ with arbitrary input obligatory location set is by no means trivial. (One just recalls the traveling salesman problem.) The pragmatical approach we use for $\boldsymbol{P}_{\text{kTrav}}$ with irregular input obligatory locations is discussed in §-7.4. However since the Nyquist grid is highly regular, it is easy to calculate sufficient fast trajectory for $\boldsymbol{P}_{\text{kTrav}}(\boldsymbol{\kappa}_{Nyq})$. (In the imaging process one has the same k space fast covering problem (§-3.1).)

From the feasible trajectory k one can get straightforwardly the feasible gradient waveform g after (4.8).

The task of RF pulse design use the calculated k space trajectory sampling points to compute the best RF waveform, a linear least squares problem.

$$(\boldsymbol{P}_{\text{RF-design}}) \begin{cases} & \boldsymbol{p} \in \mathbb{C}^N, \\ \text{Given:} & \boldsymbol{s} \in \mathbb{C}^N, \\ & \text{TOL} \\ \text{Find:} & \boldsymbol{b} \in \mathbb{C}^M \\ \text{Subject to:} & \|\boldsymbol{P} - \mathcal{S} \cdot \mathcal{T} \cdot \boldsymbol{B}\|_2 \leqslant \text{TOL} \end{cases} \tag{4.16}$$

$$\begin{aligned} \textit{Where:} \qquad & \mathcal{S} = Diag(\boldsymbol{s}) \in \mathbb{C}^{N \times N} \\ & \mathcal{T} \in \mathbb{C}^{N \times M} \text{ according to (4.11).} \end{aligned} \tag{4.17}$$

The $\boldsymbol{P}_{\textbf{RF-Design}}$ is a linear least squares problem, usually underdetermined since $\#(\boldsymbol{k}) \geqslant \#(\boldsymbol{\kappa}_{\text{nyq}})$. It can be solved standard-wisely using pseudo-inverse. The property that pseudo inverse for the underdetermined problem picks up the solution with minimal L_2-norm also matches the wish of minimizing the SAR, because L_2 norm has the physical meaning of the global energy stored in the waveform. Advanced numerical method like CG or GMRES can be applied to accelerate the computational time.

4.1.3 Two examples

In Chapter-1 an example with a heart shape profile is presented to demonstrate the conventional paradigm approaching $\boldsymbol{P}_{\text{SSEP}}$. (fig-1.3) The boundary smoothness of this profile represent roughly the requirement of the small region excitation application. The chose of this heart shape profile instead of more regular shapes like disk or rectangular is to demonstrate the independency to the symmetry of the profile content. This profile of heart shape represents a large class of profiles relevant for the application of small region excitation.

To demonstrate the capability to achieve arbitrary complex target profiles we present a second example with a much higher complexity level of the target profile, the Mona-Lisa Portait (fig-4.3-*top-left*). Of course in the real application one won't excite a Mona-Lisa profile. The aim of this example is to demonstrate the extreme case of complexe information content of the target profile.

Since the Mona-Lisa portrait is represented at the same spatial sampling grid as the previous heart shape profile, the according Nyquist grid and thus the resulting k space trajectory and gradient waveform are identical to their counterparts in the heart profile example. Therefore, despite of the large difference at the content complexity level, one ends up with the same excitation duration. The only essential difference here is the RF waveform.

The phantom experiments for the both examples were on a GE 3T MRI scanner using a ball shape water phantom (fig-7.11).

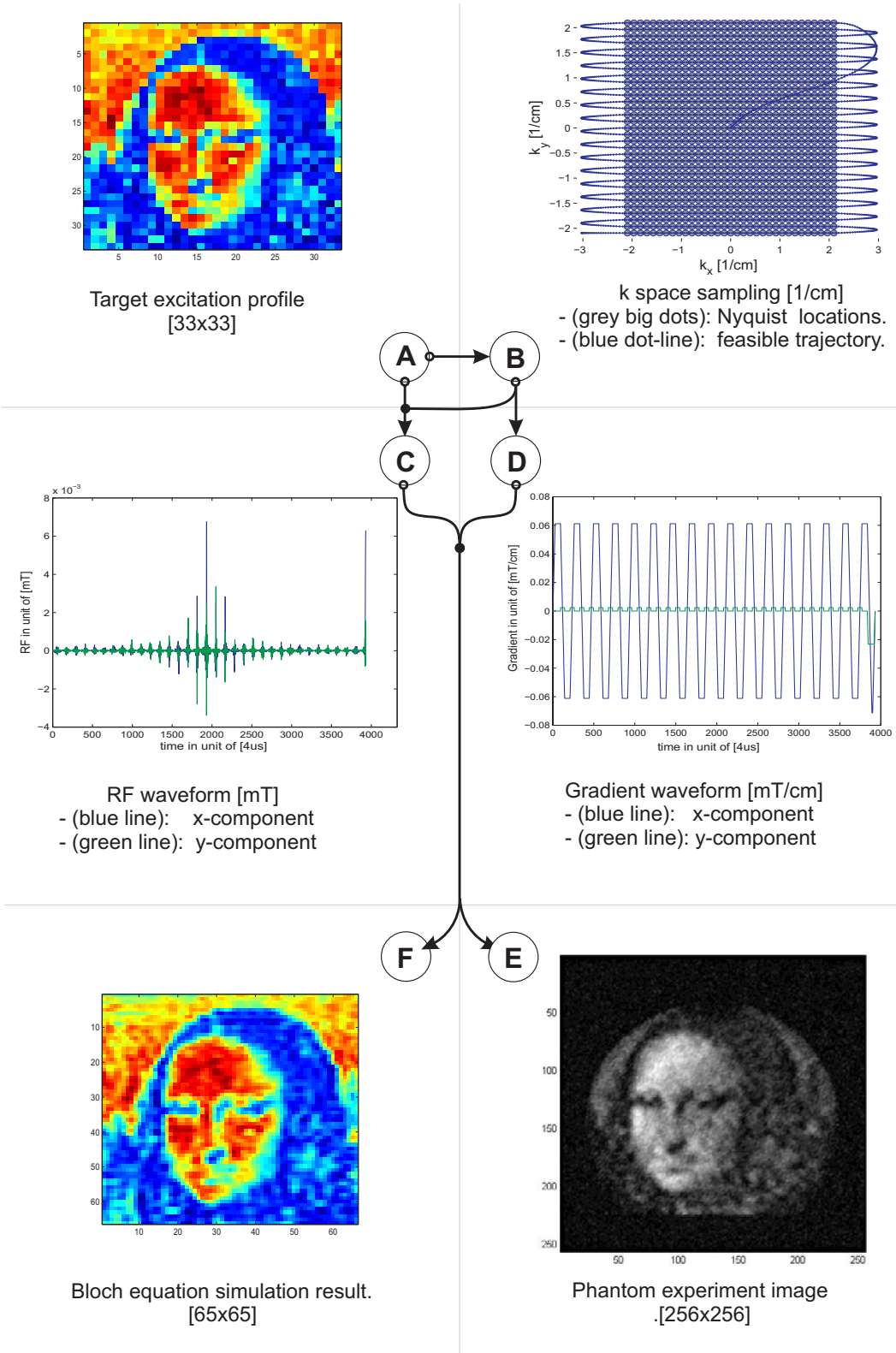


Figure 4.3: The conventional paradigma of full Nyquist STA method, Mona-Lisa Portrait

4.1.4 Accelerating the spatial selective excitation

The motivation for accelerating in the excitation process

With the introduced small tip angle method it seems one can theoretically design the excitation for any excitation target profile in no matter 1 or 2 or 3-D. But indeed there are strong restrictions. One of the major limitations is the too long excitation duration. This thesis is concentrated on this issue.

Else than in the imaging process, the motivation to accelerate the excitation is not to shorten the whole scanning process. The process duration of the excitation process is generally only a small fraction of the whole MRI scanning process, it doesn't make that much difference in the total time.

The indeed motivation to accelerate the selective excitation process comes from two aspects:

1. The model of the relaxation free Bloch equation.

   As the basis of the spatial selective excitation problem one used the simplified Bloch equation and ignored the relaxation effect with the argument that the excitation duration is much smaller than the relaxation time.

   Since the relaxation time is a fixed object property, this assumption gives effectively an upper boundary of the duration of the excitation process, which can be designed under this relaxation free model.

   This upper boundary turns to be a real restriction if one goes from 1D excitation to multi-dimensional excitation, since the according k space will also grow from 1D to 2- or 3D. The number of the according Nyquist sampling location grows from $\mathcal{O}(10^2)$ to $\mathcal{O}(10^3 \sim 10^4)$.

   A real 3D selective excitation is in general still very restricted because of this reason.

2. The model of the spatially linear varying z direction magnetic field.

   $$B_z(\xi, t) = B_0 + \langle \xi, g(t) \rangle$$

   In the model, we assume a perfect realization of B_z, the slope of which is the gradient waveform $g(t)$ as the control function.

   But in the praxis, it usually cannot be perfectly realized by the hardware. The B_0 has indeed certain spatial inhomogeneity, $B_0 \rightarrow B_0 + \delta B_0(\xi)$. The gradient term on one hand has certain non-linearity, on the other hands the slope of it has also certain error by the realization, $\langle \xi, g(t) \rangle \rightarrow \langle \xi, g(t) + \delta g(t) \rangle + \delta B_g(\xi)$.

So in general one has both spatially and temporally error in B_z:

$$B_z(\xi, t) = B_0 + \langle \xi, g(t) \rangle + \delta B_z(\xi, t)$$

The accumulation and propagation of this error δB_z is strongly correlated with the excitation duration. The shorter the excitation process, the smaller is the effect of this error. By shortening the excitation process one can largely reduce the resulting artifacts.

In MRI terminology one calls the artifact caused by the spatial inhomogeneity error of B_z the *off-resonance-artifact*. The error $\delta g(t)$ is physically largely caused by the eddy-current in the gradient coils. One call *eddy current artifact*.

Undersampling the Nyquist grid

In the STA method, the excitation duration equals the traveling duration through the calculated k space trajectory. Beside the hardware feasibility constraints, the traveling duration of the k space trajectory is largely decided by how many obligatory sampling locations it needs to cover. In the conventional approach the sampling location to be covered is the Nyquist grid. The number of the Nyquist sampling locations increases dramatically from 1D excitation profile to 2- or 3D excitation profile. A reasonable strategy for shortening the excitation process is therefore to *undersample* the Nyquist grid.

It is in general plausible that the less sampling locations need to be covered, the shorter the traveling duration would likely be. However one needs to keep in mind that beside the number of the sampling locations to cover, their distribution in the k space plays also a role on the effect of acceleration, since one still needs to solve the feasible fast covery $\boldsymbol{\kappa}$ problem.

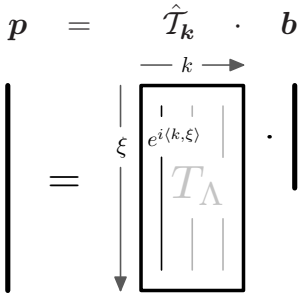
The challenge of undersampling the Nyquist sampling grid

Let's take a look at what happens if we just naively undersample the Nyquist sampling grid without any advanced treatment. What kind of issue and challenge we would have.

Recall the imaging problem, where the k space sampling ($\boldsymbol{S}$) happens in the Range of the encoding matrix (3.4). An undersampling of the signal vector $\boldsymbol{S}$ regarding to the Nyquist sampling grid causes an underdetermined the problem.

The the k space sampling of the excitation process ($\boldsymbol{b}$) happens in the Domain of the transmit matrix $\hat{\mathcal{T}}$ (4.13). In contrast to the undersampling of

the imaging process, the undersampling of the excitation process has a different nature of the challenge — an representation problem with undercomplete set of elementary vectors.



The task is to represent the magnetization vector $\boldsymbol{p} \in \mathbb{C}^N$ in the target-vector space with the column vectors of $\hat{\mathcal{T}}_{\boldsymbol{k}}$ and the coefficient vector $\boldsymbol{b} \in \mathbb{C}^M$ from the Domain of $\hat{\mathcal{T}}_{\boldsymbol{k}}$. The challenge we face here is: Due to the uncompleteness of the column vectors of the transmit matrix, there is always a subspace in the target-vector space which cannot be reasonably mapped to from the coefficient space. The uncompleteness of the column vectors is caused by the undersampling of the Nyquist grid. The size of this unreachable subspace is decided by how undercomplete the elementary vector set is. In general there would be a non-zero distance between the target vector and the representible subspace, $Range(\hat{\mathcal{T}}_{\boldsymbol{k}})$. In our excitation problem this non-zero distance will appear as the discrepancy between the target profile and the indeed achievable profile – the *excitation artifact*.

This observation leads us to the task, *how to generate a transmit matrix with its range covering the whole target-vector space.* So if we notate the target-vector space as $\mathfrak{P}$, then the target is to generate $\hat{\mathcal{T}}_{\boldsymbol{k}}$ such that

$$Range(\hat{\mathcal{T}}_{\boldsymbol{k}}) \supseteq \mathfrak{P}$$

As for the imaging process, there are obviously two alternative directions:

1. artificially expand the $Range(\hat{\mathcal{T}}_{\boldsymbol{k}})$ despite of the undersampling of the Nyquist grid;

2. try to scale down the $\mathfrak{P}$.

The currently emerging method, the parallel RF transmit approach, follows indeed the first direction – to expand the $Range(\mathcal{T}_{\boldsymbol{k}})$. The parallel transmit method can be seen as a counterpart to the parallel imaging method in the excitation side. We will introduce this approach in the next section.

In this work we will propose a now approach based on adaptive concept. It follows the second direction of scaling down the target-vector space. It is an complementary mechanism to the redundancy principle by the parallel transmit. It can be seen as having an correspondent role as the compressed sensing from the imaging process side. We will discuss the adaptive idea in §-4.3.

We recall that in the imaging process one can combine the parallel imaging approach with the compressed sensing approach maximally undersample the Nyquist grid. In this work we also combine the adaptivity principle with the redundancy principle to maximize the sparsifying capability regarding the Nyquist grid. The necessary mathematic tools for that will be introduced in the Chapter 6. The concrete implementation and validation will be introduced in the Chapter 8.

4.2 Parallel RF Transmit

This section is devoted to introduce the parallel RF transmit method, especially in the aspect of the mechanism of redundancy to enable Nyquist grid undersampling. [KBLvdB03], [Zhu04], [GYZ+06]

We will at first introduce the physical setting of parallel transmit and its intuitive motivation in §4.2.1. Then in §4.2.2, we will give the mathematical formulism of the parallel transmit problem and how it enables Nyquist grid undersampling. After that, we'll give an example to demonstrate its acceleration effect in §4.2.3.

We use the abbreviation "pTx" for *parallel Transmit*.

4.2.1 Physical setting and intuitive motivation

In the parallel transmit setting, instead of using only one RF transmit coil, one uses a series of independent RF transmit coils. Each of them has different spatial sensitivity profile depending on their coil design and positioning. The setting can be schematically described in (fig-4.4). (Compared fig-4.1)

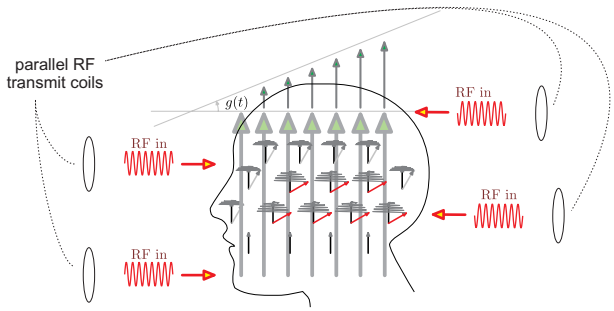


Figure 4.4: Schematic for the parallel RF transmit setting

The motivation of using parallel transmit setting used to undersample the Nyquist sampling grid can be demonstrated intuitively in the following theoretical limit case of the ideal sensitivity profiles.

For a single RF transmit coil with almost homogenous sensitivity profile like fig-4.5-*left*, the according Nyquist k grid is a regular grid with the same resolution as the target profile has, e.g. 33×33, (fig-4.5-*right*).

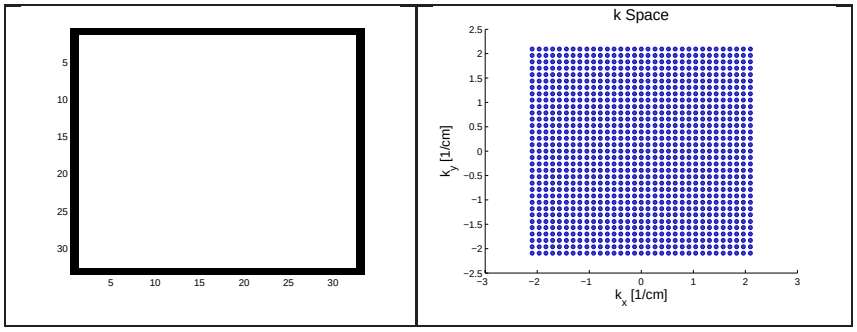


Figure 4.5: The ideal homogeneous sensitivity and its Nyquist grid (*Left:*) the ideal homogenous RF transmit sensitivity profile; (*Right:*) The according Nyquist grid.

If one has 4 independent RF coils, each of them has a ideally localized sensitivity profile as shown in fig-4.6-*left*. For each RF transmit coil the effective responsible spatial area is reduced to 1/4 of the original FOV. The according Nyquist sampling grid has then a four times larger sampling distance in k_y-direction as the original one (fig-4.5-*right*). This means a four-fold undersampling of the Nyquist k grid in the k_y direction can be allowed without any issue with the representation quality.

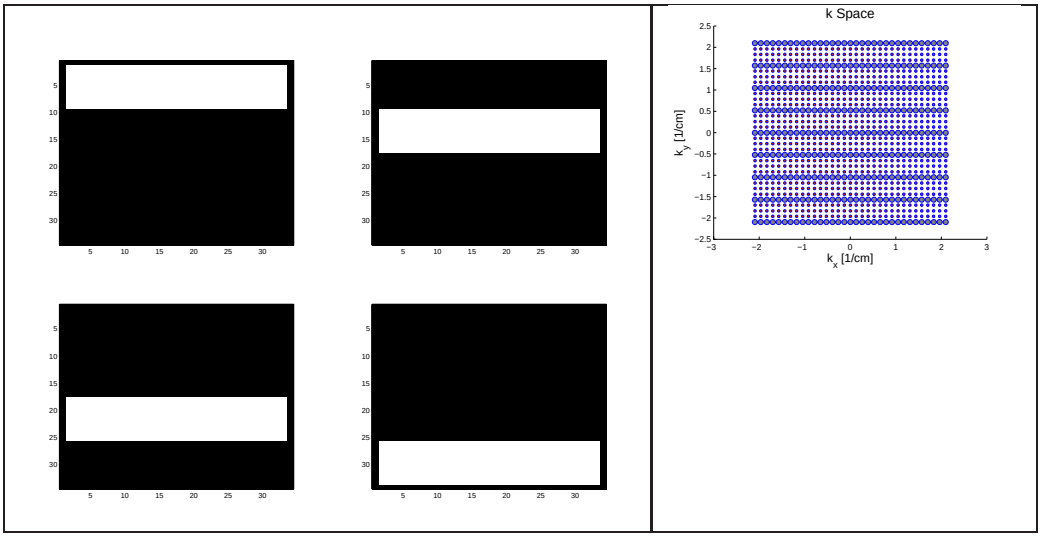


Figure 4.6: The ideal pTx sensitivity profiles and their Nyquist grid (*Left*): the ideally localized pTx sensitivity profiles; (*Right:*) Their according Nyquist grid.

Of course this is only an ideal case for intuitively demonstrating the principle of parallel RF transmit. The realistic sensitivity profile of the RF transmit coils looks different, for example in fig-4.7 are the sensitivity profiles of a real 6 channel setting. It is by far not so ideal that a six fold undersampling can be allowed.

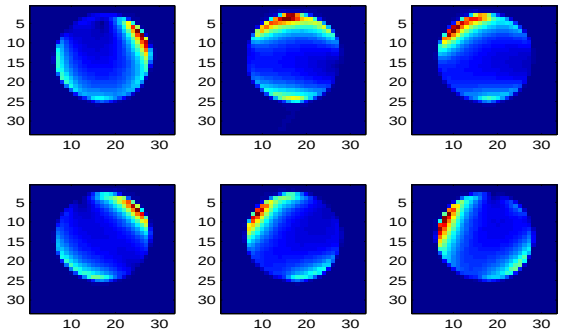
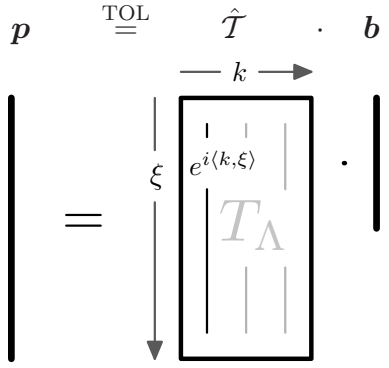


Figure 4.7: Realistic sensitivity profiles example from a 6 channel pTx setting

The pragmatic approach then is to undersample with a lower rate than the number of the independent channels. For example, one can use an undersample rate of 3∼4 with this 6 parallel channel setting. Empirically this pragmatic compromise gives satisfying result. See the example in §-4.2.3.

4.2.2 The mathematical formulism

At first we recall the naively undersampled transmit equation with respect to the Nyquist sampling grid.



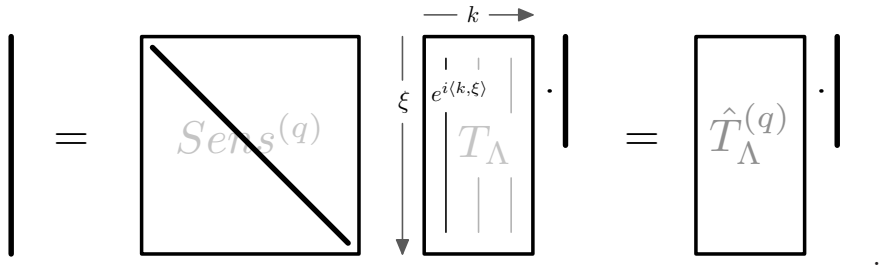
As an undercomplete representation problem, one has the issue that the column space of the transmit matrix $\hat{\mathcal{T}}$ is smaller than the target vector space $\mathfrak{P}$.

The strategy of parallel RF transmit method is trying to expand the column space of the transmit matrix $\hat{\mathcal{T}}$ despite of an undersampling of the Nyquist grid.

We remember that the general transmit matrix $\hat{\mathcal{T}}$ is the original DFT form transmit matrix $\mathcal{T}$ multiplicated by the RF spatial sensitivity matrix, $\hat{\mathcal{T}} = \mathcal{S} \cdot \mathcal{T}$. The undersampling of the Nyquist sampling grid has the effect that $\mathcal{T}$ turns to be a rectangular matrix with less column number than the row number.

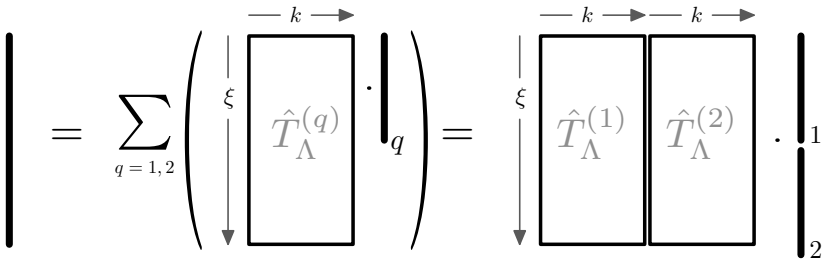
By parallel RF transmit, there are a multiple number of RF transmit coils, say C coils. Each of them has different spatial sensitivity profiles, $\mathcal{S}^{(c=1:C)}$. In the STA regime, every RF channel contribute independently to the total excitation profile. Their individual excitation process can be written as

$$\boldsymbol{M}^{(c)} = \mathcal{S}^{(c)} \cdot \mathcal{T}_\Lambda \cdot \boldsymbol{b}^{(c)} = \hat{\mathcal{T}}^{(c)} \cdot \boldsymbol{b}^{(c)} \tag{4.18}$$



Within the STA model, because of the linear nature of the transmit equation, the resulting collective excitation profile by all the channels is just the complex sum of the individually generated excitation profiles: $\boldsymbol{M} = \sum_{c=1:C} \boldsymbol{M}^{(c)}$. Then the total transmit equation, $\boldsymbol{p} \overset{\text{TOL}}{=} \boldsymbol{M}$, turns to be

$$\boldsymbol{p} \overset{\text{TOL}}{=} \sum_{c=1:C} \left(\hat{\mathcal{T}}_{\Lambda}^{(c)} \cdot \boldsymbol{b}^{(c)} \right) = \hat{\mathcal{T}}_{\Lambda} \cdot \boldsymbol{b} \tag{4.19}$$



where $c = 1 : C$

$$\boldsymbol{b} := \begin{pmatrix} \boldsymbol{b}^{(1)} \\ \vdots \\ \boldsymbol{b}^{(C)} \end{pmatrix} \in \mathbb{C}^{L \cdot C}; \qquad \boldsymbol{b}^{(c)} \in \mathbb{C}^{N} \tag{4.20}$$

$$\hat{\mathcal{T}}_{\Lambda} := \begin{pmatrix} \mathcal{T}_{\Lambda}^{(1)}, & \cdots, & \mathcal{T}_{\Lambda}^{(C)} \end{pmatrix} \in \mathbb{C}^{n \times L \cdot C}; \quad \mathcal{T}_{\Lambda}^{(c)} = (\mathcal{S}^{(c)} \cdot \mathcal{T}) \in \mathbb{C}^{n \times L}. \tag{4.21}$$

For the demonstration simplicity, the schematic is of a simple case of 2 channels setting. The principle is of course general and not restricted in the two coils case.

Depending on how different to each other the sensitivity profiles $\mathcal{S}^{(c=1:C)}$ are, the dimension column space of the total transmit matrix can be increased from the number of the k sampling locations, $|\Lambda|$. By ideal case it can be maximally increased to $C \cdot (|\Lambda|)$.

How far one can maximally undersample the Nyquist grid is decided by if the Range of the total transmit matrix $\hat{\mathcal{T}}_{\Lambda}$ can sufficiently approximate the target vector space $\mathfrak{P} = \mathbb{C}^N$.

$$\text{Range}(\hat{\mathcal{T}}_{\Lambda}) \approx \mathbb{C}^N$$

For example in the two extreme case:

1. *Optimistic limit case.* The sensitivity profiles $\mathcal{S}^{(1:C)}$ are like fig-4.6:

All column vectors of $\hat{\mathcal{T}}_\Lambda$ are still orthogonal, thus $\text{Range}(\hat{\mathcal{T}}_\Lambda) = \mathbb{C}^N$. One can thus have an undersampling rate equal the number of the coils C.

2. *Pessimistic limit case. All the $\mathcal{S}^{(1:C)}$ are identical*:

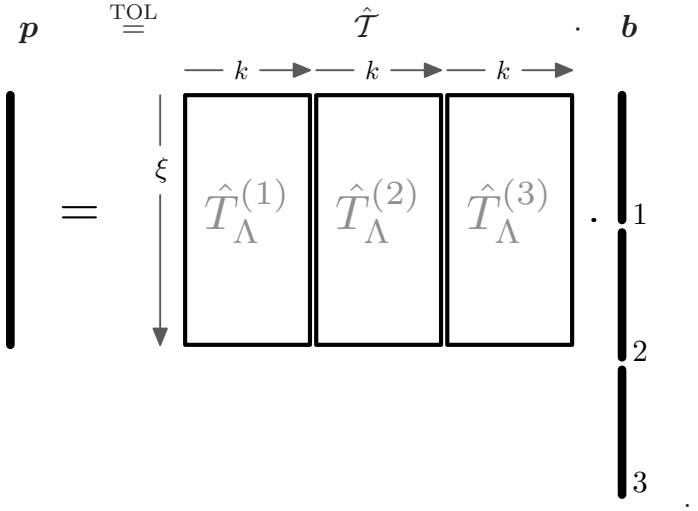
   Since $\text{Range}(\hat{\mathcal{T}}_\Lambda) = \text{Range}(\hat{\mathcal{T}}_\Lambda^{(c)}) \subset \mathbb{C}^N$, in general no undersampling is allowed.

3. *Realistic case,* For the realistic sensitivity profiles like in fig-4.7:

   Some of the column vectors of the $\hat{\mathcal{T}}_\Lambda$ are still linear dependent. If one has a undersampling rate equal the number of the coils, $\frac{|\Lambda|}{n} = C$, The range of the $\hat{\mathcal{T}}$ will not be able to cover the whole target profile space

$$\text{Range}(\hat{\mathcal{T}}_\Lambda) \subset \mathbb{C}^N.$$

   Therefore the possible undersampling rate is smaller than the number of the independent transmit channels C. For example to achieve a twice undersampling rate, one needs usually three independent RF channels:



Remember in the conventional STA approach using single RF transmit channel, the $\boldsymbol{P}_{\text{SSEP}}$ can be formulated as $\boldsymbol{P}_{\text{kTrav}}(\boldsymbol{\kappa}_{\text{Nyq}}) + \boldsymbol{P}_{\text{RF-Design}}$.

For the parallel RF transmit setting, the *parallel-*$\boldsymbol{P}_{\text{SSEP}}$ in the STA regime can similarly be formulated as

$$\boldsymbol{P}_{\text{kTrav}}(\boldsymbol{\kappa}_{\text{pTx}}) + \boldsymbol{P}_{\text{RF-Design-pTx}}$$

where $\boldsymbol{P}_{\text{kTrav}}(\boldsymbol{\kappa}_{\text{pTx}})$ is a special case of $\boldsymbol{P}_{\text{kTrav}}$ (4.15). It take the regularly undersampled Nyquist grid $\boldsymbol{\kappa}_{\text{pTx}}$ as the input obligatory locations. Since $\boldsymbol{\kappa_{pTx}}$ is usually highly regular, it is not difficult to approach.

The $\boldsymbol{P}_{\text{RF-design-pTx}}$ extend the single channel $\boldsymbol{P}_{\text{RF-design}}$ to the parallel channel setting.

$$(\boldsymbol{P}_{\text{RF-design-pTx}}) \begin{cases} & \boldsymbol{p} \in \mathbb{C}^N, \\ \text{Given:} & \boldsymbol{s} \in \mathbb{C}^N, \\ & \text{TOL} \\ \text{To find:} & \boldsymbol{b} \in \mathbb{C}^{C \cdot M} \\ \text{Subject to:} & \|\boldsymbol{p} - \hat{\mathcal{T}} \cdot \boldsymbol{b}\|_2 \leqslant \text{TOL} \end{cases} \tag{4.22}$$

$$\textit{Where:} \quad \boldsymbol{b} \in \mathbb{C}^{C \cdot M} \text{ according (4.20)}, \quad \hat{\mathcal{T}}_{\boldsymbol{k}} \in \mathbb{C}^{N \times C \cdot M} \text{ according (4.21)} \tag{4.23}$$

4.2.3 An example for parallel RF transmit excitation

We present an example of the parallel RF transmit using realistic RF sensitivity profiles as in (fig-4.7). We use the same heart shape target profile as in the example (fig-1.3). A four times undersampling in the k_y direction is applied. One still get a reasonable result despite of the undersampling.

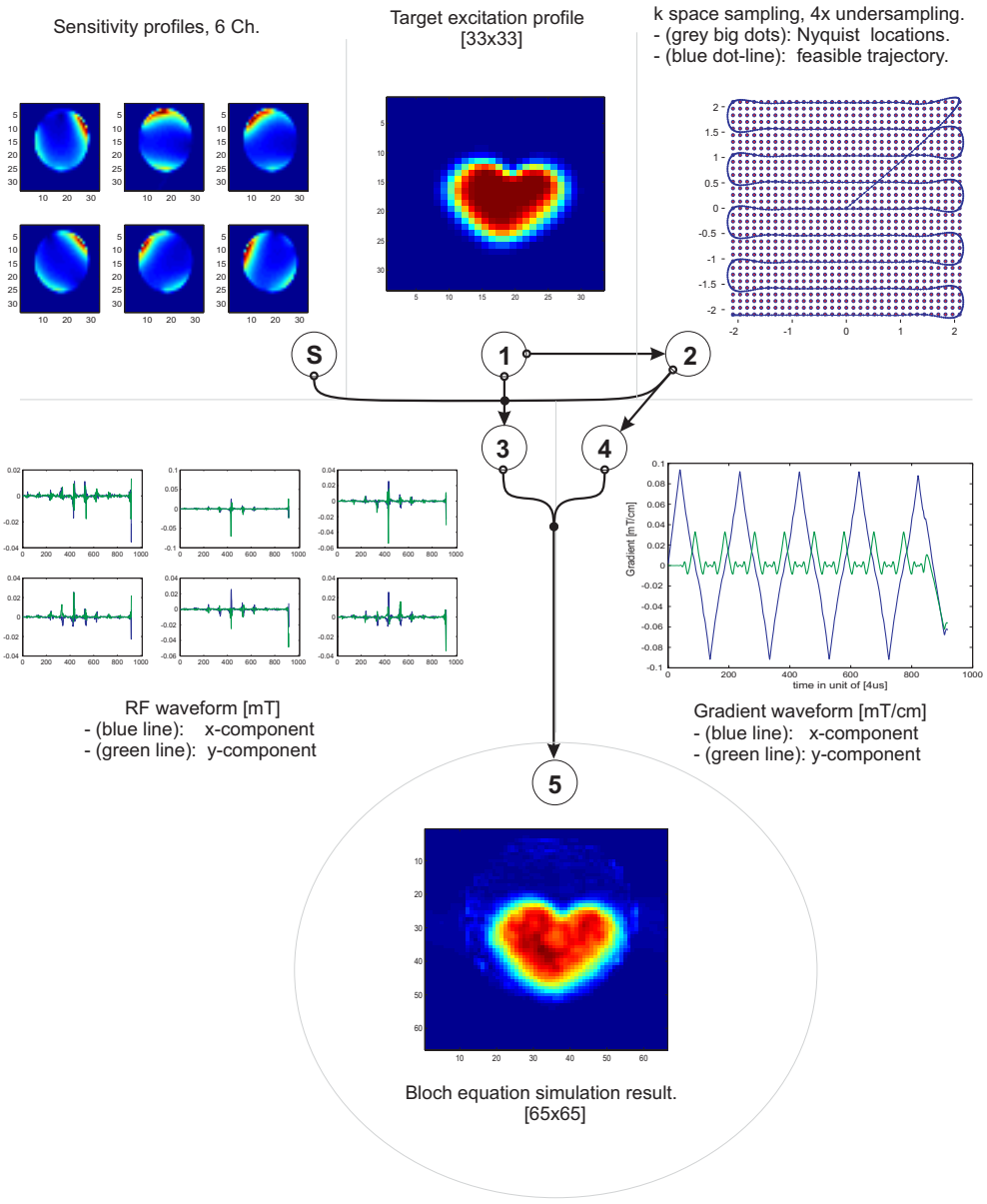


Figure 4.8: The conventional parallel transmit paradigm

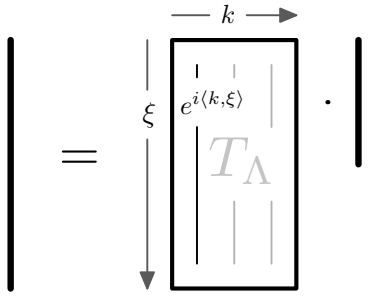
4.3 Adaptive Sparse Concept

In this section we develop the adaptive concept as an independent strategy from the parallel RF transmit concept to enable the undersampling.

4.3.1 Single Channel

We start with the undersampled transmit equation as a representation problem using undercomplete set of elementary vectors.

$$\boldsymbol{p} \overset{\text{TOL}}{=} \hat{\mathcal{T}}_\Lambda \cdot \boldsymbol{b} \tag{4.24}$$



The issue is that the Range of $\hat{\mathcal{T}}_\Lambda$ can not completely cover the target vector space $\mathfrak{P} = \mathbb{C}^N$.

$$\text{Range}(\hat{\mathcal{T}}_\Lambda) \subset \mathbb{C}^N$$

The parallel RF transmit concept follows the strategy of trying to expend the column space of $\hat{\mathcal{T}}_\Lambda$ using an undersampled frequency parameter $\boldsymbol{\kappa}$, so that

$$\text{Range}(\hat{\mathcal{T}}_\Lambda) \approx \mathbb{C}^N$$

Instead of expand the range of $\hat{\mathcal{T}}_\Lambda$, An alternative way is to analyze the problem more carefully and trying to reduce the target vector space $\mathfrak{P}$ from the whole $\mathbb{C}^N$ to a smaller subspace, so that

$$\mathfrak{P} \subset \mathbb{C}^N$$
$$\text{Range}(\hat{\mathcal{T}}_\Lambda) \approx \mathfrak{P}$$

In the imaging problem we have quite similar situation. There one wants to use the signal space to cover the domain of the encoding matrix $\hat{\mathcal{E}}_\Lambda$.

$$\mathfrak{S} \overset{!}{\approx} \text{Dom}(\hat{\mathcal{E}}_\Lambda)$$

An undersampling of the signal vector causes the problem that the signal space is smaller than the domain of $\hat{\mathcal{E}}_\Lambda$, thus can not fully cover the $\text{Dom}(\hat{\mathcal{E}}_\Lambda)$

$$\mathfrak{S} \subset \text{Dom}(\hat{\mathcal{E}}_\Lambda)$$

One way to release this issue is to expand signal space to match $\mathrm{Dom}(\hat{\mathcal{E}}_\Lambda)$, which is basically the parallel imaging approach. Another way is to reduce the domain of $\hat{\mathcal{E}}_\Lambda$ from the whole $\mathbb{C}^N$ to a smaller subspace, this is basically the idea of compressed sensing.

By compressed sensing the $\mathrm{Dom}(\hat{\mathcal{E}}_\Lambda)$ is basically the magnetization space. The key to reduce the image space is the prior-knowledge of the information sparsity of the images to be reconstructed.

In contrast to the imaging problem, where one don't know concretely the image to be reconstructed thus the information sparsity is more or less the only general prior knowledge one has. The excitation problem has the very convenient situation that one basically knows the target vector to be reconstructed. Therefore, the reduction of the target vector space from $\mathbb{C}^N$ to a smaller subspace is indeed trivial. For each concretely given target vector $\boldsymbol{p} \in \mathbb{C}^N$, the target vector space is reduced from $\mathbb{C}^N$ to the one dimensional subspace represented by this $\boldsymbol{p}$.

Now since we have an only one dimensional subspace to cover, in principle it's possible that an appropriately selected small elementary vector subset is sufficient to represent this subspace regarding the moderate user given tolerance, e.g. $\sim 5\%$. Indeed in our application it's very likely to happen because most of our target profiles has the property of strong information sparsity in the Fourier domain, which is the only relevant domain due to the physical setting.

So the principle idea of our adaptive approach is: *Tailored calculate the k space undersampling pattern after the individual given target profile* $\boldsymbol{p}$ *and user tolerance.* The concrete task is then: *To represent a given target vector using as less elementary Fourier vectors as possible.*

This task leads us to the class of the mathematical problem called sparse approximation. Here one deals with a special form of the linear representation problem. *To represent a given target vector using as less vectors as possible from a large, usually overcomplete, set of the elementary vectors.*

The major difference of the sparse approximation problem with the typical linear representation problem is (1) the requirement on the representation sparsity, (2) the use of the overcomplete set of elementary vectors. In general, introducing overcompleteness to the elementary vector set can enhance the sparsity. These two distinguishing points above causes the combinatorial nature of the sparse approximation problem. In general it's a NP hard problem.

The sparse approximation can be seen as the major underlying technique behind the compressed sensing. The compressed sensing is indeed a special extension of the sparse approximation topic with an additional part of

random sampling to preserve the low dimensional information during the undersampling.

We will introduce the general sparse approximation topic in a mathematical aspect in Chapter-5. Concretely, for the spatial selective excitation problem using a single RF transmit channel, the sparse approximation problem takes the special form that the large candidate vectors set is restricted in Fourier vectors.

Remember in the conventional STA method, one just simply chooses the complete orthogonal Fourier basis via the Shannon Nyquist theorem. The whole problem $\boldsymbol{P}_{\text{SSEP}}$ is then approached through the two subproblems:

$$\textit{Taking the Nyquist grid} + \boldsymbol{P}_{\text{kTrav}}(\boldsymbol{\kappa}_{\text{Nyq}}) + \boldsymbol{P}_{\text{RF-Design}}.$$

Now the adaptive concept changes them accordingly to the following three subproblems

$$\boldsymbol{P}_{\text{sparse}} + \boldsymbol{P}_{\text{kTrav}} + \boldsymbol{P}_{\text{RF-Design}},$$

where, $\boldsymbol{P}_{\text{sparse}}$ replaces the non-adaptive strategy of simply *taking the Nqyiust grid*. Instead of that it tailored determines an undersampling pattern as sparse as possible:

$$(\boldsymbol{P}_{\text{sparse}}) \quad \begin{cases} \text{Given:} & \begin{cases} \boldsymbol{p} \in \mathbb{C}^N, & \text{(target profile)} \\ \boldsymbol{s} \in \mathbb{C}^N, & \text{(sensitivity profile)} \\ \text{TOL} & \text{(user tolerance)} \end{cases} \\ \text{Find:} & \boldsymbol{\kappa} \in \mathbb{C}^L \text{ with } L \text{ as small as possible} \\ \text{Subject to:} & \|\boldsymbol{p} - \mathcal{S} \cdot \mathcal{T}_{\boldsymbol{\kappa}} \cdot \boldsymbol{b}\|_2 \leqslant \text{TOL} \end{cases} \tag{4.25}$$

$$\textit{Where:} \qquad \mathcal{S} = Diag(\boldsymbol{s}) \in \mathbb{C}^{N \times N}$$
$$\mathcal{T}_{\boldsymbol{k}} \in \mathbb{C}^{N \times C \cdot M} \text{ according (4.11)} \tag{4.26}$$

The second step $\boldsymbol{P}_{\text{kTrav}}$ is formally the same as in the convenional paradigm, except that now the input obligatory locations are not restricted to the regular Nyquist one. They could be quite irregular in general.

The last step $\boldsymbol{P}_{\text{RF-Design}}$ is the same as introduced in §-4.1 (4.16).

The formulation above of the adaptive sparse concept is based on the conventional setting of single RF transmit channel. We will focus on the adaptive sparse approach for the single channel setting in chapter-4.

4.3.2 Parallel Transmit

The adaptive sparse concept proposed above is independent to the parallel RF transmit concept. To maximize the sparsifying capability, it is desirable to combine the two mechanisms together.

To applying adaptive sparse undersampling while at the same time utilizing the parallel RF transmit setting can be formulated as

$$\boldsymbol{P}_{\text{sparse-pTx}} + \boldsymbol{P}_{\text{kTrav}} + \boldsymbol{P}_{\text{RF-Design-pTx}},$$

The last two subproblems $\boldsymbol{P}_{\text{kTrav}}$ and $\boldsymbol{P}_{\text{RF-Design-pTx}}$ are the same as in (4.15) and (4.22) respectively. The $\boldsymbol{P}_{\text{sparse-pTx}}$ extends the $\boldsymbol{P}_{\text{sparse}}$ to the parallel RF transmit setting.

$$
(\boldsymbol{P}_{\text{sparse-pTx}}) \quad \begin{cases} \text{Given:} & \begin{cases} \boldsymbol{p} \in \mathbb{C}^N, & \text{(target profile)} \\ \boldsymbol{s^{(c)}} \in \mathbb{C}^N, c = 1:C & \text{(sensitivity profiles)} \\ \text{TOL} & \text{(user tolerance)} \end{cases} \\ \text{Find:} & \boldsymbol{\kappa} \in \mathbb{C}^L \text{ with } L \text{ as small as possible} \\ \text{Subject to:} & \|\boldsymbol{p} - \sum_{c=1}^{C} \mathcal{S}^{(c)} \cdot \mathcal{T}_{\boldsymbol{\kappa}} \cdot \boldsymbol{b}\|_2 \leqslant \text{TOL} \end{cases}
$$

$$
\textit{Where:} \quad \begin{aligned} &\mathcal{S}^{(c)} = Diag(\boldsymbol{s^{(c)}}) \in \mathbb{C}^{N \times N}, \\ &\mathcal{T}_{\boldsymbol{k}} \in \mathbb{C}^{N \times C \cdot M} \text{ according (4.11)} \end{aligned} \tag{4.27}
$$

In the sparse approximation aspect using the parallel transmit setting to sparsely sample the k space can be seen as the benefit of getting more sparsity by introducing overcomplete dictionary.

However, compared with the conventional sparse approximation problem, the $\boldsymbol{P}_{\text{sparse-pTx}}$ has the specialness that the sparsity requirement on the sampling locations $\boldsymbol{\kappa}$ is not the same as the conventional column-sparseness of the representing matrix, $\hat{\mathcal{T}}_{\boldsymbol{\kappa}} := \left(\mathcal{S}^{(1)} \cdot \mathcal{T}_{\boldsymbol{\kappa}}, \cdots, \mathcal{S}^{(C)} \cdot \mathcal{T}_{\boldsymbol{\kappa}}\right)$. In contrast to the single channel case, where each k space sampling location correlate with single column vector in the representing matrix, in parallel transmit case each k space sampling correlate with a group of column vectors with the same frequency parameter. We call this special sparsity requirement the synchronized sparsity or subspace sparsity.

The special requirement of synchronized sparsity causes additional mathematical complexity to approach this problem. We will need to extend the

current standard numerical method and the according theoretical analysis to match this special requirement. The mathematical aspect regarding this will be discussed in the Chapter-5. The implementation and validation details of the adaptive parallel transmit will be discussed in Chapter-8.

Part II

The Math. – Sparse Approximation

Chapter 5

Sparse Approximation

In part-I we discussed the spatial selective excitation problem. Under the STA model, the goal of the sparse undersampling of the k space leads to the class of problems called *sparse approximation*.

The sparse signal approximation problem has gained more and more popularity over the last decade. A wide range of applications, such as source coding, denoising, source separation, and pattern analysis have benefited from the progress made in this area. The current chapter focuses on this topic.

The sparse approximation problem deals with the task of linear representating a given signal using a subset of vectors chosen from a large collection of elementary vectors. The number of the used vectors for the representation is should be minimized. The typical motivations for a sparse solution are, for example:

i. The number of the used vectors correlates with certain cost which one wants to control or minimize. For example, in data compression application, it correlates with the required amount of storage. In our current application of MRI, it correlates with the excitation duration, which we want to minimize.

ii. In some situations, one has the prior-knowledge of the information low-dimensionality. It means one can represent the information carried by the signal with much less data points than the formal dimensionality of the signal vector. One could to utilize this prior knowledge for eliminating unexpected errors/noise, (denoising, auto-error-correction).

Obviously, the larger the available elementary vector collection, the higher is the chance that a given target vector can be represented with very few ele-

ment vectors out of this big collection. This motivates the use of the overcomplete vector collections for sparse representation. Mallat and Zhang in their paper [MZ93] made the following comparison with the common experience in linguistics:

Writing a text with a small vocabulary, although this vocabulary might be sufficient to express all ideas, it requires to use circumvolutions that replace unavailable words by full sentence. That's why the human language vocabulary has strong redundancy — a largely overcomplete system. This overcompleteness makes the economic and elegant expression of complex thought much easier.

Compared to the classical representation problem from the elementary linear algebra, the sparse approximation using overcomplete element vector collections has two essential differences.

Sparsity: while the classical representation problem focuses basically only on how to represent the target correctly, the sparse approximation focuses on not only the representation quality but also the economic aspect of the representation, in the sense of using as few element vectors as possible.

overcompleteness: while the classical linear representation problems usually uses either complete or incomplete elementary vector sets, the sparse approximation problems usually deals with overcomplete vector collections.

These two major differences bring uncharted complexity and make the sparse approximation problem in general NP hard [DMA97], [Nat95]. To approach this NP hard sparse approximation problem, there are several plausible heuristic methods proposed, which are backed by the successful numerical experience. There is also a range of theoretical results providing the situation when rigorous optimality can be guaranteed using these proposed heuristic methods. [MZ93] [CDS99] [DETR06] [Tro04b]

Since the sparse approximation is essential technique for the implementation of our adaptive sparse concept for the selective excitation acceleration problem, the goal of this chapter is to introduce the important results in this topic.

In section 1 we introduce some terminology and the problem statement of sparse approximation.

In section 2 we introduce two major numerical approaches, the *greedy* approach and the *convex relaxation* approach. In section 3 we introduce

the theoretical results concerning the uniqueness condition for the sparse approximation problem and the optimality guarantee condition for the two numerical approaches.

As discussed in Part-I, the application of MRI parallel transmit leads to the generalization of the standard sparse approximation problem to an extended version using multi-dimensional subspaces to sparsely represent the target vector. We call it subspace sparse approximation. In the next chapter we will develop this extension in both numerical and theoretical aspects.

5.1 Definitions, Formal Problem Statement

5.1.1 Definitions and terminology

1. The *target-vector*

   The so-called *target-vector space* is an N dimensional inner-product space $\mathbb{C}$ with the usual hermitian inner-product $\langle \cdot, \cdot \rangle$ and the corresponding Euclidean norm $\| \cdot \|_2$.

   An element of the target-vector space is called *target vector* $\boldsymbol{p} \in \mathbb{C}^N$. Usually they are also called *signal* and *signal space* in the sparse approximation literature. In this work we use the name target-vector instead to avoid the potential confusion with the physically detected signal in the imaging process of MRI.

   The distance of two vectors is the Euclidean norm of their difference.

2. The *dictionary* and the *atoms*

   The large collection of candidate vectors for the purpose of sparsely representing the target vector is called *dictionary* Φ. It's a finite collection of normalized elementary vectors.

   We notate the number of the dictionary elements as Ω. If a dictionary is redundant, then $\Omega > N$.

   The elementary vectors $\varphi \in \mathbb{C}^N$ is usually called *atoms*.

   Two examples of simple dictionaries are the orthonormal Fourier basis and the Dirac basis. Both are not redundant.

   Redundant dictionary can be generated for example by concatenating several complete dictionaries. For example the time-frequency dictionary is a redundant dictionary generated by concatenating the Fourier basis together with the Dirac basis.

3. The *coefficient space*, $\mathbb{C}^\Omega$. The coefficient space $\mathbb{C}^\Omega$ consists of the coefficient vectors $\boldsymbol{b}$ representing a target vector $\boldsymbol{p} = \sum_{j=1}^{\Omega} \varphi_j \cdot \boldsymbol{b}_j$.

4. The matrix form of the dictionary, Φ

   The *Dictionary (synthesis) matrix* is defined as the matrix $\Phi \in \mathbb{C}^{N\times\Omega}$. Its column vectors are notated as φ_j. It is a mapping from the coeffient space to the target-vector space:

   $$\Phi : \mathbb{C}^\Omega \mapsto \mathbb{C}^N : \boldsymbol{b} \mapsto \sum_{j}^{\Omega} c_j \cdot \varphi_\omega$$

   The *dictionary analysis matrix* is defined as the conjugate transpose of the dictionary synthesis matrix, Φ^*. It is a mapping from the vector-space to the coefficient space:

   $$\Phi^* : \mathbb{C}^N \mapsto \mathbb{C}^\Omega : \Phi^*(\omega, :) \cdot \boldsymbol{p} = \langle \varphi_\omega, \boldsymbol{p} \rangle$$

5. The *subdictionary*, Φ_Λ

   Subdictionary means a subset of vectors of the dictionary defined by the index vector Λ. The matrix form accordingly: Φ_Λ and Φ_Λ^*.

6. The *semi l_0-norm*

   The *semi l_0-norm* is defined as the number of the non-zero entries:

   $$\|\boldsymbol{b}\|_0 = |\operatorname{supp}(\boldsymbol{b})|$$

   where *supp*($\boldsymbol{b}$) is the *support* of the vector $\boldsymbol{b}$, i.e. the set of indices whose entry is non-zero.

   The l_0-norm is NOT a norm, because doesn't have the positive scalability. Its name comes from the connection with the general p-norm: $\|\cdot\|_p \stackrel{\text{def}}{=} \left[\sum_j |c_j|^p\right]^{1/p}$, where $p > 0$.

7. the *sparsity*

   We define the *sparsity* of a linear representation of a given target vector as the number of the elementary vectors whose representation coefficients are non-zero.

   The sparsity can also be equivalently defined as the semi-l_0-norm of the coefficient vector, $\|\boldsymbol{b}\|_0$.

   Remark: The *sparsity* defined here is indeed rather a measure of the "diversity" than the "sparsity" regarding the common language. But

since no confusion would occur, we just chose this simple notation instead of at first defining the diversity as the number of non-zero entries and then defining the sparsity as its opposite, the number of the zero-entries. So, a sparser representation corresponds to the smaller zero-norm $\|\cdot\|_0$.

5.1.2 Formal problem statement

The basic problem statement of sparse approximation can be formulated as:

$$(\boldsymbol{P}_0) \quad \begin{cases} \text{Given:} & \boldsymbol{p} \in \mathbb{C}^N, \Phi \in \mathbb{C}^{N\times\Omega} \\ \min\limits_{\boldsymbol{b}\in\mathbb{C}^\Omega} \|\boldsymbol{b}\|_0, & \text{subject to } \Phi\cdot\boldsymbol{b} = \boldsymbol{p}. \end{cases} \tag{5.1}$$

$\boldsymbol{P_0}$ can be generalized to a version with a finite accuracy requirement ε:

$$(\boldsymbol{P}_{0,\varepsilon}) \quad \begin{cases} \text{Given:} & \boldsymbol{p} \in \mathbb{C}^N, \Phi \in \mathbb{C}^{N\times\Omega}, \varepsilon \in \mathbb{R}^+ \\ \min\limits_{\boldsymbol{b}\in\mathbb{C}^\Omega} \|\boldsymbol{b}\|_0, & \text{subject to } \|\Phi\cdot\boldsymbol{b} - \boldsymbol{p}\|_2 \leqslant \varepsilon. \end{cases} \tag{5.2}$$

The problem $\boldsymbol{P}_0$ can be seen as a special case of the problem $\boldsymbol{P}_{0,\varepsilon}$ with $\varepsilon = 0$ (besides the numerical error).

5.1.3 Computational complexity

While for the special case of an orthonormal (ON) dictionary the computational complexity of the sparse approximation problem is relative small, its complexity for the general non-orthogonal dictionary is NP hard. The overcomplete dictionary is obviously non-orthogonal.

For ON dictionary, the $(\boldsymbol{P}_0)$ can be reduced to a matrix-inverse problem and a successive process to find out sufficient number of basis vectors concerning given user tolerance. (See chapter 7). The matrix inverse step costs $\mathcal{O}(N^2)$. For some special matrix structure it can be further reduced, e.g. ON Fourier dictionary can use FFT with a cost of $\mathcal{O}(N\cdot\log(N))$. The successive part costs $\mathcal{O}(L\cdot N)$, where L is the number of iterations, which equals the number of the atoms with non-zero coefficient. Thus, for an ON dictionary , the numerical cost of $\boldsymbol{P_0}$ is $\mathcal{O}(N^2)$.

In case ON dictionary, the optimality regarding sparseness can always be guaranteed, (see corollary-1)

Unfortunately, for general non-orthogonal dictionary the sparse approximation problem is NP hard. One can proof that by showing its equivalence to the co-called *Exact-Cover-by-3-Sets* problem, which is a classic NP-hard problem. It can be shown that any instant of *Exact-Cover-by-3-Sets* is reducible in polynomial time to the sparse approximation problem, [DMA97] [Nat95].

5.2 Numerical Methods

In this section we present the numerical methods for the sparse approximation problem. There are basically two main approaches, the *greedy* methods and the L_1-*convex relaxation* methods.

Because of the NP hard nature of the problem, both will be at first introduced as plausible heuristic methods in this section. We will discuss their theoretical aspects in the next section.

In §-5.2.1, we introduce the so-called Matching-Pursuit (MP) method and the Orthogonal-Matching-Pursuit (OMP) method. Both methods follow the the greedy concept. In §-5.2.2 we very briefly introduce the l_1 convex relaxation method. It is also called as Basis-Pursuit(BP) method.

All methods and algorithms discussed in this section have the following inputs and outputs:

$$(\textbf{\textit{Input}}:)\quad \begin{cases} \boldsymbol{p} & \text{(the normalized target profile)} \\ \Phi = [\varphi_1, \ldots, \varphi_\Omega] & \text{(the dictionary)} \\ 0 \leqslant \mathrm{TOL} < 1 & \text{(user tolerance)} \end{cases}$$

$$(\textbf{\textit{Output}}:)\quad \begin{cases} \Lambda & \text{(the support index)} \\ \boldsymbol{b}_\Lambda \in \mathbb{C}^\Lambda & \text{(the dictionary)} \\ \textbf{Required} & \|\boldsymbol{p} - \Phi_\Lambda \cdot \boldsymbol{b}_\Lambda\| \leqslant \mathrm{TOL} \end{cases}$$

5.2.1 Greedy methods

The basic idea of greedy is to decompose the optimization problem into a sequence of sub-decision-making problems. This sequence of sub-problems will be approached in turn. One *hopes* that by making optimal decision for each sub-step the global optimal decision for the whole problem can be approximated.

At first, we introduce Matching Pursuit (MP) as a direct application of the intuitive greedy idea on the sparse approximation problem. Then we introduce the Orthogonal Matching Pursuit (OMP) as an improved version of MP.

Matching Pursuit

The basic idea of MP [MZ93] is to directly apply the greedy strategy on the sparse approximation problem. One selects the atom vectors and its coefficients one after another. In each successive step, one picks up the 'best' next atom vector by checking which vector can maximally compensate the current residual. The according coefficient of the chosen vector is its projection on the current residual. All the selected atoms and their calculated coefficients will not be changed in the later iterations.

Algorithm 1 Matching Pursuit (MP)

1: $l \leftarrow 1$, $\boldsymbol{b} \leftarrow [\,]$, $\text{res}_0 \leftarrow \boldsymbol{p}$
2: {Find the next best index}

$$\lambda_l = \arg\max_{\omega} |\langle \text{res}_{(l-1)}, \varphi_\omega \rangle|$$

3: {Update the index vector, coefficient vector}

$$\Lambda_l = (\Lambda_{l-1}, \lambda_l)$$

$$c_l = (c_{l-1}, \langle \text{res}_{(l-1)}, \varphi_\omega \rangle)$$

4: {Update the residual}

$$\text{res}_l = \text{res}_{l-1} - \langle \text{res}_{(l-1)}, \varphi_{\lambda_l} \rangle \cdot \varphi_{\lambda_l}$$

{Check for break}
5: **if** $\| \text{res}_l \|_2 > \text{TOL}$ **then**
6: $l \leftarrow l + 1$; $\rightarrow$ Step 2
7: **else**
8: **return** Λ, $\boldsymbol{b}$ with success!
9: **end if**

Numerical Cost. Let's estimate the numerical cost of the MP algorithm. Assume the greedy loop succeeds at iteration L. Step-2 requires $\mathcal{O}(\Omega \cdot N)$ operations. Step-3 and -4 require only $\mathcal{O}(N)$ operations. The total cost of the MP algorithm is then $\mathcal{O}(L \cdot \Omega \cdot N)$.

Orthogonal Matching Pursuit (OMP)

The restriction of MP The MP algorithm is intuitive and numerically efficient. However, one can easily find examples, for which the MP algorithm clearly fails.

For example Pati et. al. [Pat93] constructed a two dimensional target signal and used two non-orthogonal basis vectors as dictionary. The MP algorithm did converge. However, in general it never succeeds within 2 iterations.

The reason for that is the freezing of both the previously selected atoms and their coefficients during the later iteration steps.

In general, the loss of the optimality using MP algorithm can have two reasons. A) non-optimality due to the non global-optimally choosing of the atoms, B) non-optimality due to the non-optimally determined coefficients.

The reason A) is caused by the core combinatorial nature of the problem. But optimality loss of B) is avoidable. A refined version of MP algorithm, Orthogonal Matching Pursuit (OMP) is indeed constructed to eliminate the optimality loss due to B). It adds a least squares minimization step at each iteration to obtain the best approximation using the already selected the atoms.

Algorithm 2 Orthogonal Matching Pursuit (OMP)

1: $l \leftarrow 1$, $\boldsymbol{b} \leftarrow [\,]$, $\mathrm{res}_0 = \boldsymbol{p}$
2: {Find the next best index}

$$\lambda_l = \arg\max_{\omega} |\langle \mathrm{res}_{(l-1)}, \varphi_\omega \rangle|$$

3: {Update the index vector}

$$\Lambda_l = (\Lambda_{l-1}, \lambda_l)$$

4: {Update the coefficient vector, the residual}

$$\boldsymbol{b}_l = \underset{\boldsymbol{b} \in \mathbb{C}^l}{\operatorname{argmin}} \left(\|\boldsymbol{p} - \phi_\Lambda \cdot \boldsymbol{b}\|_2 \right)$$

$$\mathrm{res}_l := \|\boldsymbol{p} - \phi_\Lambda \cdot \boldsymbol{b}_l\|_2$$

{Check for break}
5: **if** $\| \mathrm{res}_l \|_2 > \mathrm{TOL}$ **then**
6: $l \leftarrow l + 1$; $\rightarrow$ Step 2
7: **else**
8: **return** Λ, $\boldsymbol{b}$ with success!
9: **end if**

The improvement of the OMP algorithm compared with the MP algorithm is due to the step-4. It induces the loop invariant property that the residuals are always orthogonal to the subspace spanned by the chosen atoms.

$$\langle \mathrm{res}_l, \varphi_{\lambda_i} \rangle = 0, \forall i = 1 : j$$

It follows that only the an atom that is linear independent from the already chosen atoms will be selected. As a consequence, the algorithm always succeeds at latest after N iteration steps. For instance taking the previous example, for which the MP algorithm failed, the OMP algorithm always succeeds within two steps.

Numerical Cost. The numerical cost of the OMP algorithm is determined by step-2 and step-4. The step-2 requires $\mathcal{O}(N \cdot \Omega)$ operations to perform the inner product with all the atoms in dictionary. The step-4 is a least squares task, which can be iteratively calculated based on the results from the previous step. So, within the l-th iteration loop, it costs only additional $\mathcal{O}(N \cdot l)$ operations. Totally it takes $\mathcal{O}(N \cdot L \cdot M + N \cdot L^2)$ operations. Since $L \leqslant M$, the numerical cost for OMP algorithm is:

$$\mathcal{O}(N \cdot L \cdot M).$$

5.2.2 Convex relaxation

An alternative concept to *greedy* approach is the *convex relaxation* approach.

That the sparse approximation problem is NP hard is mainly caused by the combinatorial nature of the l_0-norm. The basic idea of the convex relaxation is trying to approximate the NP hard combinatorial sparse approximation problem with a polynomial solvable convex programming problem.

The overcomplete signal representation has infinite many solutions because of the redundant nature of the dictionary. To get unique solution out of them one needs additional selection criterion, e.g. a cost function to minimize. In the case of sparse approximation problem, this additional criterion is the sparsity, which can be written as minimal-l_0. However due to the combinatorial nature of the l_0-norm, it's NP hard. The basic idea of the *convex relaxation* to approximate the NP hard combinatorial cost l_0-norm with a polynomial solvable convex cost function. So the question is to find a easier cost function to replace the l_0, but at the same time also has the property to reasonably promote the sparsest solution. The l_1-norm turns out to be a good candidate for that purpose.

The approach using l_1-minimization is also called Basis-Pursuit (BP). It basically approximates the solution of the original NP hard problem $(\boldsymbol{P_0})$ with the solution of the following convex optimization problem $(\boldsymbol{P}_1)$

$$(\boldsymbol{P}_1) \quad \begin{cases} \text{Given:} \quad \boldsymbol{p} \in \mathbb{C}^N, \Phi \in \mathbb{C}^{N \times \Omega} \\ \\ \min\limits_{\boldsymbol{b} \in \mathbb{C}^\Omega} \|\boldsymbol{b}\|_1 \quad \text{subject to } \Phi \cdot \boldsymbol{b} = \boldsymbol{p} \end{cases} \tag{5.3}$$

From the computing aspect, the l_1 minimization is a standard linear programming problem There are standard software solving it reliably and relative economically . From the sparsity promoting aspect, numerous numerical tests have shown that this approach can successfully select out the sparse solution, [CDS99].

5.3 Theoretical Analysis

In the previous two sections we discussed the sparse approximation problem and the two main numerical approaches for it, namely the greedy method and the l_1-minimization method. However, We didn't discuss the theoretical aspects like the uniqueness or the optimality. We will devote this section to these two questions:

1. *Uniqueness*

   Regarding a given dictionary, has a given target profile an unique sparsest representation?

2. *Optimal Sparsity*

   Under which condition the numerical methods can indeed find the sparsest representation?

5.3.1 Preliminary

To discuss the recent theoretical results, we at first need to introduce some preparatory concepts and definitions to further characterize the dictionaries.

Dictionary Coherence

An important property of a given dictionary is how the atoms in this dictionary differentiate from each other. The following observation may be very intuitive and plausible: Are atoms in a dictionary nearly identical to each other, then no matter how large this dictionary is, it doesn't really help to sparsely represent an arbitrary vector.

We characterize this dictionary property with the concept of *coherence*.

Definition 1 *(Mutual) coherence*

The (mutual) coherence μ *is defined as the maximal absolute inner product between two different atoms*

$$\mu := \max_{j \neq k} |\langle \varphi_j, \varphi_k \rangle|$$

Proposition 1 *([Tro04a])*

- *Obviously, every orthogonal basis has coherence zero.*

- *For general dictionary, assume $\Omega \geq N$, a lower bound of its coherence is*
$$\mu \geqslant \sqrt{\frac{\Omega - N}{N \cdot (\Omega - 1)}}$$
- *Dictionaries consisting of concatenated orthonormal bases, a so-called* multi-ONB*, has coherence lower bound $\mu \geqslant \sqrt{\frac{1}{N}}$*

The definition of the mutual coherence can be seen as measuring the similarity between any single atom with the rest in the dictionary. A refinement of it would be the measuring of the similarity between any group of the atoms, assume with m member, with the rest atoms in the dictionary. This is the motivation of the so-called *cumulative coherence*. To differentiate it with the mutual coherence μ, we notate it as μ_1:

Definition 2 *Cumulative coherence $\mu_1(m)$*
Let m be the #(atoms) in a subset of atoms from a dictionary,

$$\begin{aligned} \mu_1(m) &:= \max_{|\Lambda|=m} \max_{\omega \in \Lambda} \sum_{\lambda \in \Lambda} |\langle \varphi_\omega, \varphi_\lambda \rangle| \\ \mu_1(0) &:= 0 \end{aligned}$$

Since the mutual coherence regards only the maximal value, it reflects only the extreme case of the "likeness" between the atoms in a dictionary. However compared with the cumulative coherence, the mutual coherence has the advantage of easy to calculate.

Proposition 2 *Obviously:*

$$\begin{aligned} \mu_1(1) &= \mu \\ \mu_1(m) &\leqslant m \cdot \mu \end{aligned}$$

Spark

Another property to characterize a dictionary is the linear dependency of its atoms. For an overcomplete dictionary, all its atoms together are always linear dependent due to the overcompleteness. But there is still difference how "early" this linear dependency occurs as the #atoms increases: In some dictionaries, there is already no linear independency between the atoms of a very small subdictionary, while for some dictionaries one can still find linear independency even by a quite large subdictionary. This property is important for the uniqueness of a sparse representation.

Definition 3 *Spark*

A spark *of a dictionary* $\sigma = \text{Spark}(\Phi)$ *is the smallest possible number such that there exists a sub-group of* σ *atoms from* Φ *that are linearly dependent.*

Gram Matrix

An important instrument to study a property of a dictionary is the *Gram matrix*.

Definition 4 *Gram matrix, G*

Let $\Phi \in \mathbb{C}^{N\times L}$ *be the synthesis matrix for a given dictionary or subdictionary. We define the Gram matrix:*

$$G := \Phi^* \cdot \Phi \in \mathbb{C}^{L\times L}$$

Some properties

Gram matrix has the following properties:

- The Gram matrix is symmetric.
- If all column-vectors of Φ are linear independent, then G is *non-singular*.
- For normalized atoms, the diagonal of a G contains only ones.
- For an orthogonal dictionary, all off-diagonal entries are zero.
- For a non-orthogonal dictionary, the off-diagonal entries indicate the 'similarity' between the different atoms. The maximal absolute off-diagonal entry is the mutual coherence μ.
- Due to the definition of μ_1, one has the following boundary estimation

$$\|G\|_\infty \leqslant 1 + \mu_1(m-1)$$

Inverse Gram matrix

As a preparation for the optimality condition theorem, we want to study the upper boundary estimation of the inverse Gram matrix.

Proposition 3 *Let* G *be the Gram matrix of a (sub-)dictionary with* m *atoms, if* $\mu_1(m-1) < 1$ *then*

i. $\|G^{-1}\|_\infty = \|G^{-1}\|_1$

ii. $\|G^{-1}\|_\infty \leqslant \frac{1}{1-\mu_1(m-1)}$

Proof

i. The $\|\cdot\|_1$ and $\|\cdot\|_\infty$ norm of the Gram matrix are identical because 1) the matrix norm $\|\cdot\|_\infty$ is the maximum l_1-norm of the rows, while the $\|\cdot\|_1$ is teh maximum l_1-norm of the columns. And 2) the Gram matrix is symmetric.

ii. Now we proof the second part:

- We separate the diagonal and the off-diagonal part

$$G = I_m + A$$

The diagonal part is the identity because the dictionary is assumed to be normalized.

- From the definition of the cumulative coherence μ_1 follows

$$\|A\|_\infty = \mu_1(m-1)$$

- With the assumption $\mu_1(m-1) < 1$, we can apply the Neumann series:

$$\begin{aligned} \|G^{-1}\|_\infty &= \|\textstyle\sum_{j=0}^{\infty}(-A)^j\|_\infty \\ &\leqslant \textstyle\sum_{k=0}^{\infty}\|A\|_\infty^k \\ &= \frac{1}{1-\|A\|_\infty} \end{aligned}$$

- We replace $\|A\|_\infty$ by $\mu_1(m-1)$:

$$\|G^{-1}\|_\infty \leqslant \frac{1}{1-\mu_1(m-1)}$$

□

5.3.2 Uniqueness condition theorem

For an overcomplete dictionary, the representation of a given signal is in general *not* unique. However the solution of $(\boldsymbol{P}_0)$ could be unique, i.e. the unique sparsest solution.

Donoho et. al. found out [DETR06] *If the signal is 'sufficiently' sparse regarding the given dictionary, then it has unique sparsest representation.*

The following discussion is based on the work of Donoho et. al. [DETR06]. The concrete Lemmata structure slightly differs from the original formulation.

Theorem 1 *Uniqueness [DETR06]*
Let the dictionary $\Phi \in \mathbb{C}^{N\times\Omega}$ *have coherence* μ. *If there exists* $\boldsymbol{p} \in \mathbb{C}^N$ *and* $\boldsymbol{b} \in \mathbb{C}^\Omega$ *satisfying,*

1. $\boldsymbol{p} = \Phi \cdot \boldsymbol{b}$
2. $\|\boldsymbol{b}\|_0 < \frac{1}{2}(\frac{1}{\mu} + 1)$

then $\boldsymbol{b}$ *is the unique sparsest representation of* $\boldsymbol{p}$ *using the dictionary* Φ.

The proof outline of Theorem-1 is:

i. Connect the uniqueness with the linear dependency property of the dictionary, Spark(Φ).

ii. Connect the *linear independency* of a set of vectors with the *diagonal dominance* property of its according Gram matrix: $\sum_{j\neq i} |\langle\varphi_j, \varphi_i\rangle| < \|\varphi_i\|_2^2$.

iii. Estimate the *diagonal dominance* property of the Gram matrix through the dictionary coherence μ.

The complete Proof of theorem-1 consists of three lemmata corresponding to these three steps in the outline above.

Lemma 1 *A representation* $\boldsymbol{p} = \Phi \cdot \boldsymbol{b}$ *is necessarily the sparsest possible if* $\|\boldsymbol{b}\|_0 < \text{Spark}(\Phi)/2$
Proof
Let $m = \text{Spark}(\Phi)$. *Assume* $\boldsymbol{p}$ *having two different representations* $\boldsymbol{p} = \Phi \cdot \boldsymbol{b}_1$ *and* $\boldsymbol{p} = \Phi \cdot \boldsymbol{b}_2$ *with* $\|\boldsymbol{b}_{1,2}\| < m/2$. *One can construct a representation of the null vector* $\boldsymbol{0} = \Phi \cdot \boldsymbol{b}_3$ *with* $\boldsymbol{b}_3 = \boldsymbol{b}_1 - \boldsymbol{b}_2$.
Since the l_0*-norm obeys the triangle inequality,* $\|\boldsymbol{b}_3\|_0 \leqslant \|\boldsymbol{b}_1\|_0 + \|\boldsymbol{c_2}\|_0$, *one has* $\|\boldsymbol{c_3}\|_0 < m$, *which is a contradiction to with* $\boldsymbol{0} = \Phi \cdot \boldsymbol{b}_3$ *and the definition of the Spark as the cardinality of the smallest linear dependent subset.*
□

Lemma 2 *A set of vectors* $\varphi_{1:m}$ *are linear independent, if their according Gram matrix is diagonal dominant, i.e.*

$$\forall\, i = 1 : m,\ \sum_{j\neq i} |\langle\varphi_j, \varphi_i\rangle| > \|\varphi_i\|_2^2.$$

(if $\varphi_{1:m}$ *are normalized:* $\sum_{j\neq i} |\langle\varphi_j, \varphi_i\rangle| < 1$ *)*
Proof

a) *One property of the Gram matrix is that if $\varphi_{1:m}$ is linear independent, then their Gram matrix $G_{ij} = \{\langle\varphi_i, \varphi_j\rangle\}$ is non-singular.*

b) *Recall the term for diagonal dominance: a matrix $A = \{a_{ij}\}$ is diagonally dominant, if $|a_{ii}| > \sum_{j\neq i} |a_{ij}|$ for $\forall i$. This means for the Gram matrix: G is diagonally dominant, if $\|\varphi_i\|_2^2 > \sum_{j\neq i} |\langle\varphi_i, \varphi_j\rangle|$ for $\forall i$*

c) *(Levy-Desplanques theorem): a strictly diagonally dominant matrix is non-singular.*

The combination of a), b), c) leads to the statement.

□

Lemma 3

$$\text{Spark}(\Phi) > \frac{1}{\mu} + 1$$

Proof

a) *Let $\text{Spark}(\Phi) = m$. Any set of atoms, $\{\varphi_\Lambda\}$ with $|\Lambda| \leqslant m-1$, is linear independent. Hence any $(m-1)\times(m-1)$-leading-minor of the Gram matrix, $G^{(m-1)}$, is non-singular. According to Lemma-2 this in turn means $\|G_{i,i}^{(m-1)}\| > \sum_{j\neq i} |G_{i,j}^{(m-1)}|$ for $\forall i$. For a normalized dictionary it follows $1 > \sum_{j\neq i} |G_{i,j}^{(m-1)}|$ for $\forall i$.*

b) *The off-diagonal sum of the Gram matrix has the upper boundary the cumulative coherence μ_1. The off-diagonal sum of its m-leading-minor $G^{(m-1)}$ has the upper boundery $\sum_{j\neq i} |G_{i,j}^{(m-1)}| < \mu_1(m-1)$ for $\forall i$.*

c) *The upper boundary of $\mu_1(m)$ can be estimated using mutual coherence: $\mu_1(m-1) \leqslant (m-1)\cdot\mu$*

combine a), b), c) one get: $1 < (m-1)\cdot\mu$. Thus

$$m = \text{Spark}(\Phi) > \frac{1}{\mu} + 1$$

□

Proof of the *Theorem-1*

Due to *Lemma-3*: $\|\boldsymbol{b}\|_0 < \frac{1}{2}(\frac{1}{\mu}+1) \Rightarrow \|\boldsymbol{b}\|_0 < \frac{1}{2}\text{Spark}(\Phi)$.

Due to *Lemma-1*, c must be a unique representation, which is the claim of *Theorem-1*.

□

5.3.3 Optimality condition theorem for OMP algorithm

After we discussed the OMP and l_1-minimization methods themselves, we want to discuss, under which conditions they can find the sparsest solution.

In this subsection we discuss the optimality condition for the greedy method OMP. In the next subsection we briefly introduce the optimality condition theorem for the l_1-minimization method.

The following discussion is based on the work of J. Tropp, [Tro04b]. The concrete Lemmata structure for the proof of the optimality theorem differs slightly from the original formulation.

Theorem 2 *Optimality condition for the OMP algorithm [Tro04b]*

Let $\boldsymbol{p}$ be a target vector and Φ a dictionary with the coherence μ. If $\boldsymbol{p}$ has a representation:

$$\begin{aligned} \boldsymbol{p} &= \Phi \cdot \boldsymbol{b} \\ s.t. \quad \|\boldsymbol{b}\|_0 &< \tfrac{1}{2}(\mu^{-1} + 1), \end{aligned}$$

then the OMP method always finds the unique sparsest representation of $\boldsymbol{p}$ using the dictionary Φ.

The proof outline of *Theorem-2*:

A. Show that a sufficient condition for an exact catching of the sparsest solution using OMP method is

$$\|\Phi^*_{\text{opt}} r\| > \|\Phi^*_{\text{rest}} r\| \tag{5.4}$$

where Φ_{opt} is synthesis matrix with the optimal atoms used for sparsest representation of $\boldsymbol{p}$, Φ_{rest} is the synthesis matrix of the non-optimal atoms, r is the residual vector, $r \in \text{Span}(\Phi_{\text{opt}})$.

B. Show that $\max\limits_{\omega \in \text{rest}} \|\Phi^{\dagger}_{\text{opt}} \varphi_{\omega}\|_1 < 1$ implies (5.4). We notate $\Phi^{\dagger}_{\text{opt}}$ as the pseudo inverse of Φ_{opt}.

C. Estimate the upper boundary of $\max\limits_{\omega \in \text{rest}} \|\Phi^{\dagger}_{\text{opt}} \varphi_{\omega}\|_1$ by

$$\max_{\omega \in \text{rest}} \|\Phi^{\dagger}_{\text{opt}} \varphi_{\omega}\|_1 < 1 - \frac{1 - \mu_1(m-1) - \mu_1(m)}{1 - \mu_1(m-1)}$$

D. Combine A) B) C) one get the necessary condition for finding the sparsest representation as following:

$$\mu_1(m-1) + \mu_1(m) < 1.$$

Due to $\mu_1(m) \leqslant m \cdot \mu$, the condition formulated in μ becomes

$$m = \|\boldsymbol{p}\|_0 < \frac{1}{2}(\mu^{-1} + 1)$$

The complete proof consists of three following lemmata according to (A), (B) and (C) in the above proof-outline.

Lemma 4 *Let Φ_{opt} be the synthesis matrix of the m_{opt} optimal atoms used by the sparsest representation of $\boldsymbol{p}$. Let Φ_{rest} be the synthesis matrix of the non-optimal atoms and r be in the linear span of the optimal atoms. Then*

$$\|\Phi_{\text{opt}}^* r\| > \|\Phi_{\text{rest}}^* r\| \tag{5.5}$$

is a sufficient condition for the OMP method to find the sparsest representation.

Proof (Induction)

We show the step from n-th iteration to (n+1)-th iteration of the OMP algorithm, under the assumption that the n previously chosen atoms are all in the optimal atom set.

Since $\boldsymbol{p} \in \text{Span}(\Phi_{\text{opt}})$ and all chosen atoms are in the optimal set, the residual vector r_n must lie in the span of the optimal atoms: $r_n \in \text{Span}(\Phi_{\text{opt}})$. Since the optimal atom of current iteration step is chosen by the maximal inner-product of the atoms and the residual, condition (5.5) guarantees that only an atom from the optimal set is chosen.

After m_{opt} steps, all elements of the optimal set are chosen. The aimed optimal representation is then obtained by the least squares step.

□

Lemma 5 *For $\forall r \in \text{Span}(\Phi_{\text{opt}})$, a sufficient condition for (5.5) is*

$$\max_{\omega \in \text{rest}} \|\Phi_{\text{opt}}^{\dagger} \varphi_\omega\|_1 < 1$$

Proof

Since the ∞-norm of a matrix is its maximum of the absolute row-sum, $\|A\|_\infty = \max_i \left(\sum_j |a_{ij}|\right)$, one has the following relation:

$$\max_{\omega\in\text{rest}} \|\Phi_{\text{opt}}^\dagger \varphi_\omega\|_1 = \| \left(\Phi_{\text{opt}}^\dagger \cdot \Phi_{\text{rest}}\right)^* \|_\infty = \|\Phi_{\text{rest}}^* (\Phi_{\text{opt}}^\dagger)^* \|_\infty \tag{5.6}$$

Since $r \in \text{Span}(\Phi_{\text{opt}})$, its projection on $\text{Span}(\Phi_{\text{opt}})$ is itself. i.e. $\hat{\mathcal{P}}_{\text{opt}} \cdot r = r$. Where $\hat{\mathcal{P}}_{\text{opt}}$ notes the projection operator on $\text{Span}(\Phi_{\text{opt}})$: $\hat{\mathcal{P}}_{\text{opt}} := \Phi_{\text{opt}} \cdot \Phi_{\text{opt}}^\dagger$.

Since $\hat{\mathcal{P}}_{\text{opt}}$ is conjugate symmetric, i.e. $\hat{\mathcal{P}}_{\text{opt}} = (\Phi_{\text{opt}}^\dagger)^* \cdot \Phi_{\text{opt}}^*$,

$$\|\Phi_{\text{rest}}^* r\|_\infty = \|\Phi_{\text{rest}}^* \cdot \hat{\mathcal{P}}_{\text{opt}} \cdot r\|_\infty = \|\Phi_{\text{rest}}^* \cdot (\Phi_{\text{opt}}^\dagger)^* \cdot \Phi_{\text{opt}}^* \cdot r\|_\infty$$

With the triangle inequation of $\|\cdot\|_\infty$, one further gets

$$\|\Phi_{\text{rest}}^* r\|_\infty < \|\Phi_{\text{rest}}^* \cdot (\Phi_{\text{opt}}^\dagger)^*\|_\infty \cdot \|\Phi_{\text{opt}}^* \cdot r\|_\infty$$

Together with (6.19), the sufficient condition for (5.5) then is

$$\max_{\omega\in\text{rest}} \|\Phi_{\text{opt}}^\dagger \varphi_\omega\|_1 < 1 \tag{5.7}$$

□

Lemma 6 *let m_{opt} be the number of atoms in the optimal set, then*

$$\max_{\omega\in\text{rest}} \|\Phi_{\text{opt}}^\dagger \varphi_\omega\|_1 < 1 - \frac{1-\mu_1(m_{\text{opt}}-1) - \mu_1(m_{\text{opt}})}{1-\mu_1(m_{\text{opt}}-1)} \tag{5.8}$$

Proof

a) Due to the triangle inequation for the matrix $\|\cdot\|_1$-norm, one has:

$$\begin{aligned} \max_{\omega\in\text{rest}} \|\Phi_{\text{opt}}^\dagger \varphi_\omega\|_1 &= \max_{\omega\in\text{rest}} \|(\Phi_{\text{opt}}^* \cdot \Phi_{\text{opt}})^{-1} \cdot \Phi_{\text{opt}}^* \cdot \varphi_\omega\|_1 \\ &\leqslant \|(\Phi_{\text{opt}}^* \cdot \Phi_{\text{opt}})^{-1}\|_1 \cdot \max_{\omega\in\text{rest}} \|\Phi_{\text{opt}}^* \cdot \varphi_\omega\|_1 \end{aligned}$$

b) Due to the boundary estimation (Proposition-3) for the inverse Gram matrix $G = \Phi_{\text{opt}}^* \cdot \Phi_{\text{opt}}$, one has:

$$\|(\Phi_{\text{opt}}^* \cdot \Phi_{\text{opt}})^{-1}\|_1 = \|G_{\text{opt}}^{-1}\|_1 < \frac{1}{1-\mu_1(m_{\text{opt}}-1)}$$

c) Due to the definition of cumulative coherence μ_1, one has

$$\max_{\omega\in\text{rest}}\|\Phi^*_{\text{opt}}\cdot\varphi_\omega\|_1 = \max_{\omega\in\text{rest}}\left(\sum_{\lambda\in\text{opt}}|\langle\varphi_\omega,\varphi_\lambda\rangle|\right)\leqslant\mu_1(m_{\text{opt}})$$

One combines a), b), c) to get (6.21).

□

With the three lemmata above, Lemma-4, -5, -6, we can further complete the proof for theorem 2.

Proof for the *Theorem 2*

Combining *Lemma-4, 5* and *6*, one gets

$$\frac{1-\mu_1(m_{\text{opt}}-1)-\mu_1(m_{\text{opt}})}{1-\mu_1(m_{\text{opt}}-1)}>0$$

as a sufficient condition for a successful OMP approach finding the sparsest solution. This condition can be further reformulated to $\mu_1(m_{\text{opt}}-1)-\mu_1(m_{\text{opt}})<1$. Using the inequalty relation between the coherence and cumulative coherence, i.e.

$$\mu_1(m_{\text{opt}})\leqslant m_{\text{opt}}\cdot\mu,$$

one can express the sufficient optimality condition in μ:

$$\|\boldsymbol{b}\|_0 = m_{\text{opt}} < \tfrac{1}{2}\cdot(\mu^{-1}+1).$$

□

As a corollary, one can straightforwardly see that for an orthogonal dictionary the OMP algorithm always finds the best solution.

Corollary 1 *Sparse Optimality of the Orthogonal Dictionary*

For an orthogonal dictionary, the OMP algorithm always finds the sparsest solution.

Proof

An orthogonal dictionary has a mutual coherence zero:

$$\mu_{\text{orth}.} = 0$$

Thus $\frac{1}{2}(\mu^{-1}+1)=\infty$. The ON dictionary optimality condition follows from *theorem-2*.

□

5.3.4 The optimality condition theorem for l_1-min. method

The l_1-minimization method is a very popular method to approach the sparse approximation problem. However, for our application of fast spatial selective excitation in MRI, it turns out to be either not necessary (more detail see chapter-5, ON basis) or not suitable (more detail see chapter-6, *synchronized sparsity*).

In this work we thus concentrate on the greedy method based on OMP and present the l_1-minimization method mainly for the aim of completeness. Therefore, in this sub-section we just given the theorem itself. The proof of the theorem can be find in the paper of [DETR06]

Theorem 3 *Optimality condition for the l_1-minimization method [DETR06]*

Let $\boldsymbol{p}$ be a target vector and Φ be a dictionary with the coherence μ, if $\boldsymbol{p}$ has a representation:

$$\begin{aligned} \boldsymbol{p} &= \Phi \cdot \boldsymbol{b} \\ s.t. \quad \|\boldsymbol{b}\|_0 &< \tfrac{1}{2}(\mu^{-1} + 1), \end{aligned}$$

then the representation $\boldsymbol{b}$ is the unique solution of both $\boldsymbol{P_0}$, and $\boldsymbol{P_1}$.

This means by solving the convex optimization problem $\boldsymbol{P_1}$ one can get the solution of the combinatorial problem $\boldsymbol{P_0}$.

Chapter 6

Subspace Sparse Approximation

The current chapter is devoted to discuss the so-called *subspace sparse approximation problem*. Both numerical and theoretical aspects will be discussed.

In §-6.1 we formulate the problem of subspace sparse approximation. In §-6.2 we extend the conventional OMP method to subspace level to approach the subspace sparse approximation problem. We call the extended version the subspace OMP method. In §-6.3 we focus on the theoretical aspects of the subspace sparse approximation. The theoretical analysis for the subspace sparse approximation is structured in an analog fashion as the theory section (§-5.3) of the conventional sparse approximation problem.

6.1 Subspace Sparse Approximation

Sometimes the special application circumstances leads to special requirements on the sparsity. In this section we generalize the conventional sparse approximation problem regarding such a modified sparsity requirement. We call it synchronized sparsity or subspace sparsity. We call the modified problem statement the subspace sparse approximation. This problem is raised from the effort to combine the adaptive sparse excitation concept with the parallel transmit setting in MRI (See chapter 8).

The *Subspace Sparse Approximation* Problem

Let $\{v_j \in \mathbb{C}^N | j = 1 : \Omega\}$ be a collection of elementary vectors. The vectors are grouped into equal-sized small groups

$$\mathfrak{s}_k := \{v_{\lambda_k^c} | c = 1 : C;\ \lambda_k^c \in \{1 : \Omega\}\},\ k = 1 : M, \Omega = MC.$$

Let $\boldsymbol{p} \in \mathbb{C}^N$ be a target vector. The task of subspace sparse approximation is to represent the target vector using a minimal number of groups.

The conventional sparse approximation problem can be seen as a special case of $C = 1$.

We easily see, that the standard approaches for the conventional sparse approximation problem can not be directly applied for the subspace sparse approximation problem. Because they promote as few participating elementary vectors as possible, while the subspace sparse approximation problem requires as few participating groups as possible.

In general, directly applying the approaches like OMP or l_1-minimization leads to the situation that many groups drop a part of their member vectors. But since not all the members are eliminated, one cannot disregard the whole group, which is the objective.

To release this issue, we take OMP approach as starting point and modify it accordingly to match the special sparsity requirement. The next subsection introduce the modified OMP algorithm.

To prepare the further discussion, we at first extend the definition of atoms and dictionary to *subspace atom* and *subspace dictionary*.

Definition 5 *Subspace Atom, Subspace Dictionary*

Let $\{v_j \in \mathbb{C}^N | j = 1 : \Omega\}$ be a collection of elementary vectors. Where $\Omega = M \cdot C, M \in \mathbb{N}, C \in \mathbb{N}$. The grouping index vector $\lambda_k^c \in \{1 : \Omega\}$ orders the elementary vectors into a sub-group structure with M groups, each group has C members:

$$\{\{v_{\lambda_k}^c \in \mathbb{C}^N\} | k = 1 : M, c = 1 : C\}$$

The subspace atom $\mathfrak{s}_k$ *is defined as the linear span of the vector members of the group with index $k = 1 : M$.*

$$\mathfrak{s}_k := \mathrm{Span}(v_{\lambda_k^1}, \ldots, v_{\lambda_k^C})$$

The subspace dictionary *is defined as the collection of the M subspace atoms.*

$$\mathfrak{D} := \{\mathfrak{s}_k, k = 1 : M\}$$

The subspace atoms are linear spaces and the subspace dictionaries are set of linear spaces. We use the fraktur font to notate them respectively. For the discussion convenience we also define the matrix form of both. To distinguish the subspace atoms and dictionary with their matrix form, we notate the matrix forms with a hat above, $\hat{\mathfrak{s}}$, $\hat{\mathfrak{D}}$.

Definition 6 *The matrix form of subspace-atom and subspace-dictionary*

The subspace-atom matrix $\hat{\mathfrak{s}}$ is defined as a $N \times C$ matrix. Its column vectors are the members of the vector group with index k:

$$\hat{\mathfrak{s}}_k := (v_{\lambda_k(1)}, \ldots, v_{\lambda_k(C)}) \in \mathbb{C}^{N \times C}.$$

The subspace-dictionary matrix $\hat{\mathfrak{D}}$ is defined as a $N \times M \cdot C$ matrix, column-wise concatenated by all the subspace-atom matrix $\hat{\mathfrak{s}}$:

$$\hat{\mathfrak{D}} := (\hat{\mathfrak{s}}_1, \ldots, \hat{\mathfrak{s}}_M) \in \mathbb{C}^{N \times CM}.$$

Definition 7 *Subspace-subdictionary and its matrix form*

Similar to vector-atom subdictionary, we define the subspace subdictionary as a subset of $\mathfrak{D}$. The subset is notated with an index vector Λ, which picks out L values from $1 : M$:

$$\mathfrak{D}_\Lambda := \{\mathfrak{s}_{\Lambda(1)}, \ldots, \mathfrak{s}_{\Lambda(L)}\}; \Lambda(j) \in \{1 : M\}, j = 1 : L$$

Sometimes we also directly notate it as $\mathfrak{s}_\Lambda$ or $\mathfrak{s}_{\boldsymbol{k}}$.

Accordingly, its matrix form is

$$\hat{\mathfrak{D}}_\Lambda := \left(\hat{\mathfrak{s}}_{\Lambda(1)}, \ldots, \hat{\mathfrak{s}}_{\Lambda(L)}\right) \in \mathbb{C}^{N \times CL}.$$

6.2 The Subspace-OMP Algorithm

The special requirement on group sparsity leads to a modified OMP algorithm: the so-called *subspace-OMP* algorithm.

The basic idea. The basic idea of the subspace OMP algorithm is the following:

1. We take the greedy spirit of the OMP, but instead of compare and select the elementary vectors in each iteration we compare and select the whole group of the elementary vectors.

2. The OMP weights the importance of each elementary vector by looking at its "similarity" with the current residual. It picks up the most "similar" one to compensate the residual in each iteration.

   We can extend this "comparing similiarity" concept to our goal of comparing not only the single elementary vectors but the whole group of vectors.

The "similarity" of a elementary vector to the residual vector can be measured by the absolute value of their inner product, which is the projection of the residual vector on the subspace spanned by the elementary vector.

The "similarity" of a group of vectors to the residual vector can be measured by the euclidean norm of the projection of the residual vector on the multi-dimensional subspace spanned by all the member vectors in the group.

The idea above straightforwardly induced the subspace-OMP algorithm. The algorithm has the following input and output parameters.

$$(\boldsymbol{Input:}) \quad \begin{cases} \boldsymbol{p} \in \mathbb{C}^N & \text{(the normalized target profile)} \\ \{v_j \in \mathbb{C}^N | j = 1:\Omega, \Omega = C \cdot M\} & \text{(the collection of elementary vectors)} \\ \lambda_k^c \in \{1:\Omega\}, k = 1:M, c = 1:C & \text{(grouping index)} \\ 0 \leqslant \mathrm{TOL} < 1 & \text{(user tolerance)} \end{cases}$$

$$(\boldsymbol{Output:}) \quad \begin{cases} \Lambda & \text{(the support index)} \\ \boldsymbol{b}_\Lambda \in \mathbb{C}^\Lambda & \text{(the coefficient vector)} \\ \textbf{Required} & \|\boldsymbol{p} - \hat{\mathfrak{D}}_\Lambda \cdot \boldsymbol{b}_\Lambda\| \leqslant \mathrm{TOL} \end{cases}$$

The symbole $\hat{\mathcal{P}}_{\mathfrak{s}_k}(v)$ stands for the projection operator. It gives the projection of the vector v on the subspace $\mathfrak{s}_k$.

Algorithm 3 Subspace OMP

1: {Initial.}

$$l \leftarrow 1, \quad \Lambda \leftarrow [\,], \quad \text{res}_0 \leftarrow \boldsymbol{p}$$

2: {Find the next best index}

$$\lambda_l = \underset{k}{\text{argmax}} \left(\|\hat{\mathcal{P}}_{\boldsymbol{s}_k}(\text{res}_{l-1})\|_2 \right)$$

3: {Update the index vector}

$$\Lambda_l = (\Lambda_{l-1}, \lambda_l)$$

4: {Update the coefficient vector, the residual}

$$\boldsymbol{b}_l = \underset{\boldsymbol{b} \in \mathbb{C}^l}{\text{argmin}} \left(\|\boldsymbol{p} - \phi_\Lambda \cdot \boldsymbol{b}\|_2 \right)$$

$$\text{res}_l := \|\boldsymbol{p} - \phi_\Lambda \cdot \boldsymbol{b}_l\|_2$$

{Check for break}

5: **if** $\| \text{res}_l \|_2 > \text{TOL}$ **then**

6: $l \leftarrow l + 1$; $\rightarrow$ Step 2

7: **else**

8: **return** Λ, $\boldsymbol{b}$ with success!

9: **end if**

Compared with the OMP algorithm (§-5.2.1), the subspace OMP only differs in step-2. Here one extends the "maximal inner-product" selection criterion by the maximal of the projections on the multi-dimensional subspace atoms.

The computational cost. The dominant computational cost of the subspace OMP algorithm is caused by step-1. For each iteration loop, a $N \times C$ least squares problem for each candidate group has to be solved. Each least squares problem has the computational cost $\mathcal{O}(N \cdot C^2)$. With M the total number of the candidate groups, the cost of step-1 in each iteration is $\mathcal{O}(M \cdot N \cdot C^2)$.

The remaining steps, basically step-3, have the same numerical cost as in the OMP algorithm. It's about $\mathcal{O}(N \cdot L)$, which is an order of magnitude smaller than the cost of step-1.

Totally, one ends up with a computational cost of $\mathcal{O}(L \cdot M \cdot N \cdot C^2)$, where L is the number of iterations.

This algorithm is supported by successfull numerical tests, (chapter-5). Beside the heuristical success via numerical tests, we also want to give a theoretical analysis of the uniqueness and the optimality condition, which is the topic of the next section.

6.3 Theoretical analysis

6.3.1 Preliminary

Before we give the theorem for the optimality condition and its proof, we at first do some preparatory work extending the essential concepts for the vector-atoms and -dictionaries to subspace-atoms and -dictionaries.

The Concepts Extension

In §-6.1 we defined the subspace atom, subspace dictionary (def.-5). In this section we generalize the conventional transpose operation for their according synthesis matrix form (def.-8). We will also define the appropriate vector and operator norm for the coefficient space.

For a conventional matrix D the conjugate transpose D^* is indeed a projection of the input vector to each of the column-vector of D and the output vector is their projection coefficients. In the case of a subspace dictionary matrix, the appropriate extension is to project the input vector onto all its subspace atoms rather than the directly projection onto their underlying column vectors. This consideration leads us to the following definition:

Definition 8 *Generalized transpose for a subspace dictionary matrix*
Let $\hat{\mathfrak{D}} := (\hat{\mathfrak{s}}_1, \ldots, \hat{\mathfrak{s}}_M) \in \mathbb{C}^{N \times CM}$ *be a subspace dictionary matrix. Its conjugate transposed matrix is defined as*

$$\hat{\mathfrak{D}}^{\hat{*}} := \begin{pmatrix} \hat{\mathfrak{s}}_1^{\dagger} \\ \vdots \\ \hat{\mathfrak{s}}_M^{\dagger} \end{pmatrix} \in \mathbb{C}^{CM \times N}$$

where $\hat{\mathfrak{s}}^{\dagger} := (\hat{\mathfrak{s}}^* \hat{\mathfrak{s}})^{-1} \hat{\mathfrak{s}}^* \in \mathbb{C}^{C \times N}$ *is the Moore-Penrose pseudoinverse of the subspace matrix* $\hat{\mathfrak{s}}$*.*

At this point it's not obvious whether the new defined transposed matrix $\hat{\mathfrak{D}}^{\hat{*}}$ has the same matrix multiplication property like conventional matrix,

e.g. association rule etc. However, for orthogonal subspace atoms matrix the generalized transpose of their dictionary matrix is identical with the normal transpose matrix. We will discuss the orthogonalization of the subspace atoms in the next subsection.

Definition 9 *Coefficient Space, Target Space*

The target-vector space is a finite-dimensional, complex linear space $\mathbb{C}^N$. It has standard definitions of multiplication-with-scalars, addition, and inner-product.

The representation of a target vector via the a set of C dimensional subspaces is parametrized by the coefficients regarding each subspaces. Each coeffient of a subspace is itself a C-dimensional vector, $u_j \in \mathbb{C}^C, j = 1 : M$. We concatenate them all into one single column vector:

$$u \;:=\; \begin{pmatrix} u_1 \\ \vdots \\ u_M \end{pmatrix} \in \mathbb{C}^{C\cdot M\times 1}, \quad u_{j=1:M} \in \mathbb{C}^{C\times 1} \tag{6.1}$$

We call it coefficient vector of the regarding set of subspaces. We call the associated space the coefficient space.

The coefficient space has standard definition of multiplication-with-scalars and addition. However the inner-product is different. The inner-product is indeed a 1D special case of the matrix conjugate transpose and a matrix vector multiplication. We have a generalized definition of the conjugate tranpose. Hence the vector inner-product and the norm in the coefficient space must be accordingly re-defined.

Definition 10 *The inner product and the vector norm of the coefficient space. We define the inner product of the coefficient space as:*

$$\langle v, u\rangle = \hat{v^*} \cdot u. \tag{6.2}$$

We define the vector norm of coefficient space as:

$$\|u\|_{q|p} \;:=\; \left\| \begin{pmatrix} \|u_1\|_q \\ \vdots \\ \|u_{_M}\|_q \end{pmatrix} \right\|_p . \tag{6.3}$$

where $\|\cdot\|_p$ and $\|\cdot\|_q$ are the Hölder norm

One can easily see that $\|\cdot\|_{2|2}$-norm is induced by the inner-product (6.2). One can also see that this definition makes sense regarding step-2 in the **algorithm 5** (Subspace-OMP) .

Lemma 7 *Let u be the coefficient vector defined as 6.1, let $\|u\|_{q|p}$ be defined as in (6.3). The $\|u\|_{q|p}$ is a norm in the sense of*

1. $u \neq 0 \Rightarrow \|u\|_{q|p} > 0$; $u = 0 \Rightarrow \|u\|_{q|p} = 0$
2. $a \in \mathbb{C}$, $\|a \cdot u\|_{q|p} = |a| \cdot \|u\|_{q|p}$
3. $\|u^{(1)} + u^{(2)}\|_{q|p} \leqslant \|u^{(1)}\|_{q|p} + \|u^{(2)}\|_{q|p}$

Proof

Let $w := \begin{pmatrix} w_1 \\ \vdots \\ w_M \end{pmatrix}$, $w_j = \|u_j\|_q$, then $\|u\|_{q|p} = \|w\|_p$.

The hoelder norm fulfills the three norm criteria above, i.e.

1. $x \neq 0 \;\Rightarrow\; \|x\|_p > 0$
2. $a \in \mathbb{C}$, $\|a \cdot x\|_p = |a| \cdot \|x\|_p$
3. $\|x_1 + x_2\|_p \leqslant \|x_1\|_p + \|x_2\|_p$

By applying these three points on the new defined $\|\cdot\|_{q|p}$ norm, one can verify their according properties:

1. $u \neq 0, \;\Rightarrow\; \exists j, w_j = \|u_j\|_q > 0, \;\Rightarrow\; w \neq 0 \;\Rightarrow\; \|u\|_{q|p} = \|w\|_p > 0$
2. $a \in \mathbb{C}, \forall j, \|a \cdot u_j\|_q = |a| \cdot w_j = |a| \cdot \|u_j\|_q,$
   $\Rightarrow\; \|a \cdot u\|_{q|p} = \||a| \cdot w\|_p = |a| \cdot \|w\|_p = |a| \cdot \|u\|_{q|p}$
3. Let $w^{(1)} := \begin{pmatrix} \|u_1^{(1)}\|_q \\ \vdots \\ \|u_M^{(1)}\|_q \end{pmatrix}$, $w^{(2)} := \begin{pmatrix} \|u_1^{(2)}\|_q \\ \vdots \\ \|u_M^{(2)}\|_q \end{pmatrix}$, $w^{(+)} := \begin{pmatrix} \|u_1^{(1)} + u_1^{(2)}\|_q \\ \vdots \\ \|u_M^{(1)} + u_M^{(2)}\|_q \end{pmatrix}$

Then $w^{(+)} \preccurlyeq (w^{(1)} + w^{(2)})$. ("$\preccurlyeq$" means componenten-wise).

Therefore:

$$\begin{aligned} \|u^{(1)} + u^{(2)}\|_{q|p} = \|w^{(+)}\|_p &\leqslant \|w^{(1)} + w^{(2)}\|_p \\ &\leqslant \|w^{(1)}\|_p + \|w^{(2)}\|_p = \|u^{(1)}\|_{q|p} + \|u^{(2)}\|_{q|p} \end{aligned}$$

□

Since we have now two spaces: the target vector space with a euclidian vector norm and the coefficient space with the norm defined as in (6.3). We should distringuish the four types of linear operators accordingly: the operators between the signal spaces, between the coefficient spaces, from coefficient space to signal space, and from signal space to coefficient space. For the purpose of this work, we are mostly interested in the linear operator between the coefficient spaces.

Definition 11 *Linear operators between the coefficient spaces*

Let $\boldsymbol{A}$ be a linear operator between the coefficient spaces. $\mathcal{A}$ can be writen as block matrix:

$$\boldsymbol{A} := \begin{pmatrix} \boldsymbol{\alpha}_{1,1} & \cdots & \boldsymbol{\alpha}_{1,M} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\alpha}_{M,1} & \cdots & \boldsymbol{\alpha}_{M,M} \end{pmatrix} \in \mathbb{C}^{MC \times MC} \tag{6.4}$$

where $\boldsymbol{\alpha}_{i,j} \in \mathbb{C}^{C \times C}$.

We define the norm for the linear operators between coefficient spaces according to the vector norm $\|\cdot\|_{q|p}$:

Definition 12 *Norm for the linear operators between the coefficient space*

Let $\boldsymbol{A}$ be a matrix form of the operator between coefficient spaces as defined in (6.4), then $\|A\|_{q|p}$ is defined as:

$$\|\boldsymbol{A}\|_{q|p} := \| \begin{pmatrix} \|\boldsymbol{\alpha}_{1,1}\|_q & \cdots & \|\boldsymbol{\alpha}_{1,M}\|_q \\ \vdots & \ddots & \vdots \\ \|\boldsymbol{\alpha}_{M,1}\|_q & \cdots & \|\boldsymbol{\alpha}_{M,M}\|_q \end{pmatrix} \|_p$$

where $\|\boldsymbol{\alpha}_{i,j}\|_q$ is the matrix (q,q)-Hölder norm for $\boldsymbol{\alpha}_{i,j} \in \mathbb{C}^{C \times C}$, and $\|\cdot\|_p$ is the matrix (p,p)-Hölder norm.

Analog to the proof of Lemma-7, one can show that this definition fulfills also the three (vector) norm requirements.

What remains to show is, whether this definition provides a *consistent* norm concerning matrix multiplication.

Lemma 8 *The matrix norm* $\|\cdot\|_{q|p}$ *is consistent in the sense of*

$$\|\boldsymbol{A}\cdot\boldsymbol{B}\|_{q|p} \leqslant \|\boldsymbol{A}\|_{q|p}\cdot\|\boldsymbol{B}\|_{q|p}$$

Proof

Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be defined as in (6.4). The letters $\boldsymbol{\alpha}_{i,j}$, $\boldsymbol{\beta}_{i,j}$ ... denote their sub-block matrices.

We introduce the following notation rules:

- For $\boldsymbol{\alpha}_{i,j}$, the i,j are the indexes for the sub-blocks, not the element of the sub-blocks, for which we write $\boldsymbol{\alpha}_{i,j}(k,l)$.
- *Define:* $a_{i,j} := \|\boldsymbol{\alpha}_{i,j}\|_q \in \mathbb{R}$, $b_{i,j} := \|\boldsymbol{\beta}_{i,j}\|_q \in \mathbb{C}$,
  $A := (a_{i,j}) \in \mathbb{R}^{N\times H}$, $B := (b_{i,j}) \in \mathbb{R}^{H\times M}$
  $\Rightarrow \|\boldsymbol{A}\|_{q|p} = \|A\|_p$, $\|\boldsymbol{B}\|_{q|p} = \|B\|_p$
- The compact notation of sum (Einstein's convention):

$$\boldsymbol{\alpha}_i^j\cdot\boldsymbol{\beta}_j^k := \sum_{j=1}^{H}\boldsymbol{\alpha}_{i,j}\cdot\boldsymbol{\beta}_{j,k}$$

For $\boldsymbol{A}\in\mathbb{C}^{QN\times QH}$, $\boldsymbol{B}\in\mathbb{C}^{QH\times QM}$, $j = 1:C\cdot H$, one can write

$$\boldsymbol{A}\cdot\boldsymbol{B} = \begin{pmatrix} \boldsymbol{\alpha}_1^j\boldsymbol{\beta}_j^1 & \cdots & \boldsymbol{\alpha}_1^j\boldsymbol{\beta}_j^M \\ \vdots & \ddots & \vdots \\ \boldsymbol{\alpha}_N^j\boldsymbol{\beta}_j^1 & \cdots & \boldsymbol{\alpha}_N^j\boldsymbol{\beta}_j^M \end{pmatrix}$$

Since the normal matrix Hölder norm are consistent in the sense of $\|A\cdot B\|_p \leqslant \|A\|_p\cdot\|B\|_p$, we have

$$\begin{aligned} & \|\boldsymbol{\alpha}_{i,j}\cdot\boldsymbol{\beta}_{j,k}\|_q \leqslant \|\boldsymbol{\alpha}_{i,j}\|_q\cdot\|\boldsymbol{\beta}_{j,k}\|_q \\ \Rightarrow \quad & \|\boldsymbol{\alpha}_i^j\cdot\boldsymbol{\beta}_j^k\|_q \leqslant \|\boldsymbol{\alpha}_i^j\|_q\cdot\|\boldsymbol{\beta}_j^k\|_q \qquad \forall i = 1:N, \forall j = 1:M \end{aligned}$$

(The compact notation of sum is applied in the last row)

Since

$$0 \preccurlyeq \begin{pmatrix} \|\boldsymbol{\alpha}_1^j \boldsymbol{\beta}_j^1\|_q & \cdots & \|\boldsymbol{\alpha}_1^j \boldsymbol{\beta}_j^M\|_q \\ \vdots & \ddots & \vdots \\ \|\boldsymbol{\alpha}_N^j \boldsymbol{\beta}_j^1\|_q & \cdots & \|\boldsymbol{\alpha}_N^j \boldsymbol{\beta}_j^M\|_q \end{pmatrix} \preccurlyeq \begin{pmatrix} \|\boldsymbol{\alpha}_1^j\|_q \cdot \|\boldsymbol{\beta}_j^1\|_q & \cdots & \|\boldsymbol{\alpha}_1^j\|_q \cdot \|\boldsymbol{\beta}_j^M\|_q \\ \vdots & \ddots & \vdots \\ \|\boldsymbol{\alpha}_N^j\| \cdot \|\boldsymbol{\beta}_j^1\|_q & \cdots & \|\boldsymbol{\alpha}_N^j\|_q \cdot \|\boldsymbol{\beta}_j^M\|_q \end{pmatrix}$$

where "$\preccurlyeq$" stands for component-wise "$\leqslant$", we have

$$\begin{aligned} \|\boldsymbol{A} \cdot \boldsymbol{B}\|_{q|p} &= \left\| \begin{pmatrix} \|\boldsymbol{\alpha}_1^j \boldsymbol{\beta}_j^1\|_q & \cdots & \|\boldsymbol{\alpha}_1^j \boldsymbol{\beta}_j^M\|_q \\ \vdots & \ddots & \vdots \\ \|\boldsymbol{\alpha}_N^j \boldsymbol{\beta}_j^1\|_q & \cdots & \|\boldsymbol{\alpha}_N^j \boldsymbol{\beta}_j^M\|_q \end{pmatrix} \right\|_p \\ &\leqslant \left\| \begin{pmatrix} \|\boldsymbol{\alpha}_1^j\|_q \cdot \|\boldsymbol{\beta}_j^1\|_q & \cdots & \|\boldsymbol{\alpha}_1^j\|_q \cdot \|\boldsymbol{\beta}_j^M\|_q \\ \vdots & \ddots & \vdots \\ \|\boldsymbol{\alpha}_N^j\| \cdot \|\boldsymbol{\beta}_j^1\|_q & \cdots & \|\boldsymbol{\alpha}_N^j\|_q \cdot \|\boldsymbol{\beta}_j^M\|_q \end{pmatrix} \right\|_p \\ &= \left\| \begin{pmatrix} \|\boldsymbol{\alpha}_1^j\|_q & \cdots & \|\boldsymbol{\alpha}_1^j\|_q \\ \vdots & \ddots & \vdots \\ \|\boldsymbol{\alpha}_N^j\|_q & \cdots & \|\boldsymbol{\alpha}_{N\|_q}^j \end{pmatrix} \cdot \begin{pmatrix} \|\boldsymbol{\beta}_1^j\|_q & \cdots & \|\boldsymbol{\beta}_1^j\|_q \\ \vdots & \ddots & \vdots \\ \|\boldsymbol{\beta}_N^j\|_q & \cdots & \|\boldsymbol{\beta}_{N\|_q}^j \end{pmatrix} \right\|_p \\ &= \|A \cdot B\|_p \\ &\leqslant \|A\|_p \cdot \|B\|_p \\ &= \|\boldsymbol{A}\|_{q|p} \cdot \|\boldsymbol{B}\|_{q|p} \end{aligned}$$

□

The *Normalization* of the Subspace Atom

There are two reasons for introducing the concept of *"normalization"* for the subspace dictionary:

- the analogy to normal vector dictionary, where one can restricts the investigation normalized atom vectors without loss of generality to avoid the unnecessary complication due to the scaling of atoms.
- the technical difficulty to treat the generalized transpose for a subspace dictionary and the induced special 'matrix' multiplication, inner-product concept, and the further derivative concepts like pseudo inverse, etc. We will see that with the orthogonalized subspace atoms

the generalized transpose reduced to normal transpose. The "orthogonalization" of the subspace atoms is an extension of the "normalization" of the conventional vector atoms.

The core idea of a normalized conventional dictionary is the standarization of the dictionary. In this way one can keep only the characteristic information of a dictionary, which are its atom directions and shift the unnecessary information — the scaling of the atoms — out to the coefficient vector. We will apply the same concept to the subspace dictionary/atoms.

To represent a C-dimensional subspace one need C linear independent vectors of this subspace. But this representation is of course not unique. The different representations of the same subspace bring in distracting information, which have nothing to do with the characterization of the subspaces. Thus we want to eliminate these distractions by standarizing the subspace representation. One straightforward option is the orthogonalization.

By performing a QR-decomposition of an arbitrary representation $A \in \mathbb{C}^{N\times C}$ of a subspace, one can get an orthonormal representation $C \in \mathbb{C}^{N\times C}$ of the subspace. While the coefficient for a conventional vector atom is a scalar, the generalized "coefficient" of the subspace is a C-dimensional vector. The upper matrix $R \in \mathbb{C}^{C\times C}$ can then be seen as generalized scaling factor for a multi-dimensional subspace. If the subspace is one dimensinoal, the whole concept reduces to the conventional normalization of the vector atoms, where both the coefficient and the scaling factor are scalar.

With this QR-normalization concept one can restrict the investigation to the case of orthogonal representation of the subspaces without loss of generality. Any non-orthogonal representation can be reduced to it by multiplying the according scaling matrix R to its C-dimensional coefficient.

The orthogonal matrix representation of a subspace has a very nice property:

$$\hat{\mathfrak{s}}^{\dagger} = \hat{\mathfrak{s}}^{*}$$

Hence the generalized transpose for a subspace dictionary turns out to be the same as a conventional transpose. Therefore it obeys the same matrix multiplication rules. We can thus use all the derivative concepts based on matrix transpose such as the inner-product, the pseudo-inverse etc as usual. This property is essential for the later derivation of both, uniqueness and optimality theorems.

In the rest of this thesis, all the matrix representations of the subspace atoms are assumed to be orthogonal.

The dictionary coherence quantity under the subspace aspect

Due to the "QR-normalization" of the subspace atoms, the subspace dictionary has formally the same Gram matrix as for the according conventional dictionary with the same elementary vectors.

However, the difference lies in the treatment of its block strucure, defined as in (6.4), and its quantitative "measure" using the $\|\cdot\|_{q|p}$ norm.

$$\begin{aligned} G &= \hat{\mathfrak{D}}^* \cdot \hat{\mathfrak{D}} \in \mathbb{C}^{CM \times CM} \\ &= \begin{pmatrix} \hat{\mathfrak{s}}_1^* \hat{\mathfrak{s}}_1 & \cdots & \hat{\mathfrak{s}}_1^* \hat{\mathfrak{s}}_M \\ \vdots & \ddots & \vdots \\ \hat{\mathfrak{s}}_M^* \hat{\mathfrak{s}}_1 & \cdots & \hat{\mathfrak{s}}_M^* \hat{\mathfrak{s}}_M \end{pmatrix}. \end{aligned}$$

We introduce the notation: $\mathfrak{g}_{i,j} := \hat{\mathfrak{s}}_i^* \hat{\mathfrak{s}}_j \in \mathbb{C}^{C \times C}$

$$G = \begin{pmatrix} \mathfrak{g}_{1,1} & \cdots & \mathfrak{g}_{1,M} \\ \vdots & \ddots & \vdots \\ \mathfrak{g}_{M,1} & \cdots & \mathfrak{g}_{M,M} \end{pmatrix} \tag{6.5}$$

Since $\hat{\mathfrak{s}}_k$ is column-orthonormal matrix, the diagonal blocks of the Gram matrix are identity matrices: $\mathfrak{g}_{i,i} = \mathbb{1}_C$, $\|\mathfrak{g}_{i,i}\|_2 = \sqrt{C}$.

The special "measure" via the $\|\cdot\|_{q|p}$ norm induces a generalized definition of dictionary mutual coherence and the cumulative coherence.

Definition 13 *(Generalized) mutual coherence for a subspace dictionary*

The (generalized) mutual coherence for a subspace dictionary is defined as

$$\mu^{(C)} := \max_{i \neq j} (\|\mathfrak{g}_{i,j}\|_2)$$

The superscript $\cdot^{(C)}$ is to denote the dimension of the subspace: $\mathfrak{g}_{i,j} \in \mathbb{C}^{C \times C}$. Obviously, $\mu^{(1)} = \mu$

Definition 14 *(Generalized) cumulative coherence*

Let $1 \leq m \leq M$, the cumulative coherence regarding m is defined as:

$$\mu_1^{(C)}(m) := \max_{|\Lambda|=M} \max_{w \notin \Lambda} \sum_{\lambda \in \Lambda} \|\mathfrak{g}_{\omega,\lambda}\|_2$$

Obviously, $\mu_1^{(1)} = \mu_1$

Lemma 9 *Estimating $\mu_1^{(C)}$ by $\mu^{(C)}$:*

$$\mu_1^{(C)}(m) < M \cdot \mu^{(C)}$$

Proof the Lemma follows directly from the definitions.

Lemma 10 *Estimating the upper bound of the inversed Gram matrix*
For $\mu_1^{(C)}(m-1) < \sqrt{C}$ the inverse of the sub Gram matrix $\|G_\Lambda^{-1}\|_{2|\infty}$, $|\Lambda| = m$, has the following upper bound:

$$\|G_\Lambda^{-1}\|_{2|\infty} = \|G_\Lambda^{-1}\|_{2|1} \leqslant \frac{1}{1 - \mu_1^{(C)}(m-1)} \tag{6.6}$$

Proof

One can proof this estimation exactly the same way as for the estimation of $\|G^{-1}\|_\infty$, *Proposition-3*. Here is the outline:

- Seperate the diagonal and the off-diagonal parts: $G = I_M + A$.
- $\|A\|_{2|\infty} = \mu_1^{(C)}(m-1)$, per definition of $\mu_1^{(C)}$.
- For $\mu_1^{(C)}(m-1) < \sqrt{C}$, the Neumann series of $(\|\mathfrak{g}_{i,i}\|_2)$ converges

$$\begin{aligned} \|G_\Lambda\|_{2|\infty} &= \|\textstyle\sum_k^\infty (-A)^k\|_{2|\infty} \\ &\leqslant \textstyle\sum_k^\infty \|(-A)^k\|_{2|\infty} \\ &= (1 - ||A\|_{2|\infty})^{-1} \end{aligned}$$

- Replace $\|A\|_{2|\infty}$ with $\mu_1^{(C)}(m-1)$ yielding the estimation (6.6)

□

One should be aware that this estimation is only justified under the assumption of $(\mu_1^{(C)}(m-1) < \sqrt{C})$.

Now we extend the concept of linear (in)dependency to subspaces.

Definition 15 *Subspace-linear-independency*
A set of subspaces $\{\mathfrak{s}_i | i = 1 : L\}$ is linear independent, if no element subspace can be represented by the others:

$$\forall \xi_{j=1:L} \in \mathbb{C}^C : \ \textstyle\sum_{j \neq i} \hat{\mathfrak{s}}_j \cdot \xi_j \neq \hat{\mathfrak{s}}_i \cdot \xi_i, \forall i = 1 : L \tag{6.7}$$

The linear dependent set of subspaces is defined as the opposite

$$\exists \xi_{j=1:L} \in \mathbb{C}^C : \ \textstyle\sum_{j \neq i} \hat{\mathfrak{s}}_j \cdot \xi_j = \hat{\mathfrak{s}}_i \cdot \xi_i, \exists i = 1 : L \tag{6.8}$$

Recall that in the vector atom dictionary case one defines the *Spark* as the smallst number of the linear dependent atoms from the given dictionary. We extend this spark concept accordingly to the subspace dictionary.

Definition 16 *Subspace-dictionary-spark* $\mathrm{SSpark}(\mathfrak{D})$

$\mathrm{SSpark}(\mathfrak{D})$ *is the cardinality of the smallest possible linear dependent subset of a subspace-dicitonary* $\mathfrak{D}$*.*

Lemma 11 *Given a set of subspaces* $\mathfrak{D}_\Lambda := \{\mathfrak{s}_\Lambda | \Lambda = 1 : L\}$ *and the according Gram matrix* $G^\Lambda = \hat{\mathfrak{D}}_\Lambda^* \cdot \hat{\mathfrak{D}}_\Lambda$*. If the Gram matrix is block-wise strictly diagonally dominant in the sense of*

$$\forall i \in \{1 : L\} : \|\mathfrak{g}_{i,i}^\Lambda\|_2 > \sum_j \|\mathfrak{g}_{i,j}^\Lambda\|_2,$$

then this set of subspace is subspace-linear-independent in the sense of (6.7) in (def. 15).

Proof

We show that if a subspace set is subspace-linear-dependent, its Gram matrix can*not* be block-wise diagonally dominant.

Let $\{\mathfrak{s}_0, \mathfrak{s}_1, \ldots, \mathfrak{s}_L\}$ be linear dependent in the sense of (6.8)

$$\exists \xi_{0:L} \in \mathbb{C}^C \quad \textstyle\sum_{j=1:L} \hat{\mathfrak{s}}_j \cdot \xi_j = \hat{\mathfrak{s}}_0 \cdot \xi_0 \tag{6.9}$$

Without loss of generality, one can re-number the index such that

$$\|\xi_0\|_2 < \|\xi_j\|_2, \forall j = 1 : L \tag{6.10}$$

We introduce the following notation

$$\begin{aligned}
\hat{\mathfrak{D}}'_\Lambda &:= && (\hat{\mathfrak{s}}_1, \dots, \hat{\mathfrak{s}}_L) \in \mathbb{C}^{N \times CL} \\
\hat{\mathfrak{D}}_\Lambda &:= && (\hat{\mathfrak{D}}'_\Lambda, \hat{\mathfrak{s}}_0) \in \mathbb{C}^{N \times C(L+1)} \\
G_\Lambda &:= && \hat{\mathfrak{D}}_\Lambda \cdot \hat{\mathfrak{D}}^*_\Lambda \in \mathbb{C}^{C(L+1) \times C(L+1)} \\
g_0 &:= && \begin{pmatrix} \mathfrak{s}_0^* \cdot \mathfrak{s}_0 \\ \mathfrak{s}_1^* \cdot \mathfrak{s}_0 \\ \vdots \\ \mathfrak{s}_L^* \cdot \mathfrak{s}_0 \end{pmatrix} \in \mathbb{C}^{C(L+1) \times 1} \\
\mathbb{C}^{C(L+1) \times CL} \ni F_\Lambda : & && G_\Lambda =: (F_\Lambda, g_0) \\
x &:= && \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_L \end{pmatrix} \in \mathbb{C}^{CL \times 1}
\end{aligned}$$

By multiplying $\hat{\mathfrak{D}}'^*_\Lambda$ on both side of (6.9) one get

$$F_\Lambda \cdot x = g_0 \cdot \xi_0$$

The first block row of F_Λ reads:

$$\tilde{g}_0 := (\hat{\mathfrak{s}}_1^* \cdot \hat{\mathfrak{s}}_0, \dots, \hat{\mathfrak{s}}_L^* \cdot \hat{\mathfrak{s}}_0) = (\mathfrak{g}_{1,0}, \dots, \mathfrak{g}_{L,0}) \in \mathbb{C}^{C \times L}$$

Hence:

$$\sum_{j=1:L} \mathfrak{g}_{j,0} \cdot \xi_j = \mathfrak{g}_{0,0} \cdot \xi_0$$

$\Rightarrow$

$$\| \sum_{j=1:L} \mathfrak{g}_{j,0} \cdot \xi_j \|_2 = \| \mathfrak{g}_{0,0} \cdot \xi_0 \|_2$$

Since all $\mathfrak{s}$ are orthonormal, all $\mathfrak{g}_{i,i}$ are identity matrices. Thus the right hand side reads

$$\| \mathfrak{g}_{0,0} \cdot \xi_0 \|_2 = \| \mathfrak{g}_{0,0} \|_2 \cdot \| \xi_0 \|_2 = \sqrt{C} \cdot \| \xi_0 \|_2.$$

The left hand side is bounded:

$$\| \sum_{j=1:L} \mathfrak{g}_{j,0} \cdot \xi_j \|_2 \leqslant \sum_{j=1:L} \| \mathfrak{g}_{j,0} \cdot \xi_j \|_2 \leqslant \sum_{j=1:L} \| \mathfrak{g}_{j,0} \|_2 \cdot \| \xi_j \|_2$$

Since ξ_0 has been chosen chosen to have the minimal l_2 norm (6.10), we get the situation of *non* diagonal dominance

$$\sum_{j=1:L} \|\mathfrak{g}_{j,0}\|_2 \geqslant \|\mathfrak{g}_{0,0}\|_2 = \sqrt{C}$$

So, subspace-linear-dependency implies block-wise non-diagonally-dominance of their Gram matrix, which means equivalently: *The block-wise strictly diagonally dominance implies subspace linear independency.*

□

Lemma 12

$$\text{SSpark}(\mathfrak{D}) > \frac{C}{\mu^{(C)}} + 1$$

Proof

i. Since $\mu_1^{(C)}(m-1) \leqslant (m-1) \cdot \mu^{(C)}$ (Lemma 9), we get

$$\mu_1^{(C)}(m-1) < \sqrt{C}.$$

So, $M < \frac{\sqrt{C}}{\mu_1^{(C)}} + 1$ implies $\mu_1^{(C)}(m-1) < \sqrt{C}$.

ii. *To show:* $\mu_1^{(C)}(m-1) < \sqrt{C}$ implies the blockwise strict diagonally dominance of the according Gram matrix:

$$\mu_1^{(C)}(m-1) = \max_i \sum_j \|\mathfrak{g}_{i,j}^{\Lambda}\|_2 \geqslant \sum_j \|\mathfrak{g}_{i,j}^{\Lambda}\|_2, \forall i$$

Since the $\hat{\mathfrak{s}}$ are column-wise orthonormal, $\|\mathfrak{g}_{i,i}\|_2 = \sqrt{C}$. Together with $\mu_1^{(C)}(m-1) < \sqrt{C}$ one gets

$$\begin{aligned} \forall i = \{1:m\}: \quad & \textstyle\sum_j \|\mathfrak{g}_{i,j}^{\Lambda}\|_2 \leqslant \mu_1^{(C)}(m-1) < \sqrt{C} = \|\mathfrak{g}_{i,i}\|_2 \\ \Rightarrow \quad & \textstyle\sum_j \|\mathfrak{g}_{i,j}^{\Lambda}\|_2 < \|\mathfrak{g}_{i,i}\|_2. \end{aligned}$$

This is the blockwise strictly diagonally dominance.

iii. Together with lemma-11 one get the following statement:

Given a subspace dictinoary $\mathfrak{D}$ with a mutual coherence $\mu^{(C)}$, *any* m-member-subdictionary with $m < \frac{\sqrt{C}}{\mu_1^{(C)}} + 1$ is subspace linear independent, which means $\text{SSpark}(\mathfrak{D}) > m$. Thus

$$\text{SSpark}(\mathfrak{D}) > \frac{\sqrt{C}}{\mu_1^{(C)}} + 1$$

□

6.3.2 Uniqueness condition theorem

Preparations

We defined the subspace linear independency in (*Definition 15*). We can also formulate this definition in a dimensionality aspect.

Definition 17 *Subspace linear independency (dimensionality aspect)*
A set of subspaces $\{\mathfrak{s}_i \in \mathbb{C}^C | i = 1 : M\}$ *is linear independent, if its matrix representation has a kernel with the dimension smaller than C.*

$$dim\left(\ker(\hat{\mathfrak{s}}_1, \ldots, \hat{\mathfrak{s}}_M)\right) < C$$

It's obvious that the two definitions (*Definition 15*) and (*Definition 17*) are equivalent.

Definition 18 *Subspaces representating subspace*
A subspace $\hat{\mathfrak{p}} \in \mathbb{C}^{N \times C}$ *can be* represented *by the subspace dictionary* $\mathfrak{D} = \{\mathfrak{s}_{\Lambda(1)}, ...\mathfrak{s}_{\Lambda(L)}\}$ *means*

$$\forall \xi_0 \exists \xi_{1:L} \sum_{j=1:L} \hat{\mathfrak{s}}_{\Lambda_1(j)} \cdot \xi_j = \hat{\mathfrak{p}} \cdot \xi_0$$

We call $\hat{\mathfrak{p}}$ *the* target subspace*.*

Lemma 13 *Subspace-subspace uniqueness*
Let $\hat{\mathfrak{p}} \in \mathbb{C}^{N \times C}$ *represent the target subspace. The C column vectors of* $\hat{\mathfrak{p}}$ *are linear independent. Let* $\mathfrak{D}$ *be the subspace dictionary.*
If $\hat{\mathfrak{p}}$ *can be represented by a number of subspaces smaller than* $\frac{1}{2}\,\mathrm{SSpark}(\mathfrak{D})$*, Then this representation is the unique sparsest representation for this subspace.*

Proof
We assume there exist two representations of $\hat{\mathfrak{p}}$ with the same number of participating subspaces. Let two index vectors Λ_1, Λ_2 parameterize the representations, $|\Lambda_1| = |\Lambda_2| = L$:

$$\begin{array}{ll} \forall \xi_0 \exists \xi_{1:L} & \sum_{j=1:L} \hat{\mathfrak{s}}_{\Lambda_1(j)} \cdot \xi_j = \hat{\mathfrak{p}} \cdot \xi_0 \\ \forall \beta_0 \exists \beta_{1:L} & \sum_{i=1:L} \hat{\mathfrak{s}}_{\Lambda_2(i)} \cdot \beta_i = \hat{\mathfrak{p}} \cdot \beta_0. \end{array}$$

This means the intersection of $\mathrm{Im}(\hat{\mathfrak{s}}_{\Lambda_1})$ and $\mathrm{Im}(\hat{\mathfrak{s}}_{\Lambda_2})$ is at least as large as $\mathfrak{p}$.

$$\mathfrak{p} \subset (\mathrm{Im}(\mathfrak{s}_{\Lambda_1}) \cap \mathrm{Im}(\mathfrak{s}_{\Lambda_2}))$$

which means

$$\dim(\mathrm{Im}(\mathfrak{s}_{\Lambda_1}) \cap \mathrm{Im}(\mathfrak{s}_{\Lambda_2})) \geqslant C$$

We notate the union index vector $\Lambda := \Lambda_1 \cup \Lambda_2$. Further we get

$$\dim(\ker(\mathfrak{s}_{\Lambda})) \geqslant C$$

Which in turn means the set of subspaces $\{\mathfrak{s}_{\Lambda}\}$ is linear dependent. At the same time one has $|\Lambda| \leqslant |\Lambda_1| + |\Lambda_2| = 2L < SSpark(\mathfrak{D})$. Recall $SSpark(\mathfrak{D})$ is per definition the smallest number of the subspace elements, which can form a linear dependent subset. So it's contradictory to the claim: $|\Lambda| < \frac{1}{2}SSpark(\mathfrak{D})$ *and* $\{\mathfrak{s}_{\Lambda}\}$ linear dependent.

So the representation of the subspace $\mathfrak{p} \in \mathbb{C}^{N \times C}$ with less than $\frac{1}{2}M$ subspace elements must be unique.

□

Untill now the analysis is very analog to the proof for the uniqueness of represention using vector atoms, [DETR06]. The difference is that lemma 13 deals with the situation using subspaces to represent subspaces. However, we want to deal with the situation using subspace to represent a vector instead of a whole subspace.

Compared with the subspace-target-subspace-atoms case, the problematic part of the vector-target-subspace-atoms case is that there could be a non-zero intersection of the linear independent subspaces. For those target vector, which lie in this intersection area, it could happen that some subspaces are equal-valued, which would destroy the uniqueness.

This situation will not happen for the vector-target-vector-atoms case because all intersections between linear independent atoms are zero. For the subspace-target-subspace-atoms case, this situation will also not happen because of the definition of subspace-linear-independency.

Hence we want to restrict our statement to the non-intersection region. At first we want to characterize the intersection region.

Definition 19 *Intersection region H*

Given a set of subspaces $\{\mathfrak{s}_{\Lambda(1)}, ..., \mathfrak{s}_{\Lambda(L)}\}$*:*

$$H(\mathfrak{s}_{\Lambda}) \quad := \bigcup_{i=1:L} \left(\mathfrak{s}_i \cap \mathrm{Span}_{j \neq i}(\mathfrak{s}_j) \right)$$

Definition 20 *SLI Intersection Region (SLI stands for Subspace-Linear-Independent)*

Given a subspace dictionary $\mathfrak{D} = \{\mathfrak{s}_1, \cdots, \mathfrak{s}_M\}$*:*

$$\hat{H}(\mathfrak{D}) := \bigcup_{\forall\ SLI\ subset:\ \Lambda_{SLI}} (H(\mathfrak{s}_{\Lambda_{SLI}})) \tag{6.11}$$

The following (*Lemma 14*) and (*Corollary 2*) will show that the intersection region is a zero-set in $\mathbb{C}^N$.

Lemma 14 *If a set of subspaces* $\{\mathfrak{s}_{\Lambda(1)}, ..., \mathfrak{s}_{\Lambda(L)}\}$ *is subspace linear independent as Def.-15, then*

$$\begin{aligned} H(\mathfrak{s}_\Lambda) &\subset \mathrm{Span}(\hat{\mathfrak{s}}_\Lambda) \\ \dim(H(\mathfrak{s}_\Lambda)) &< \dim(\mathrm{Span}(\hat{\mathfrak{s}}_\Lambda)) \end{aligned}$$

Proof

Since subspace-set $\{\mathfrak{s}_\Lambda\}$ is subspace-linear independent, then for any index vector $|\Lambda_2| < L = |\Lambda|$:

$$\mathrm{Span}(\hat{\mathfrak{s}}_\Lambda) \subset \mathrm{Span}(\hat{\mathfrak{s}}_\Lambda).$$

It means for $\forall i$

$$(\mathfrak{s}_i \cap \mathrm{Span}(\hat{\mathfrak{s}}_j | j = \{1:L\} \backslash \{i\})) \subset \mathrm{Span}(\hat{\mathfrak{s}}_\Lambda)$$

Since L is finite, we have

$$H(\mathfrak{s}_\Lambda) := \bigcup_{i \in \Lambda} \left(\mathfrak{s}_i \cap \mathrm{Span}(\hat{\mathfrak{s}}_{\Lambda \backslash \{i\}})\right) \subset \mathrm{Span}(\hat{\mathfrak{s}}_\Lambda)$$

□

Corollary 2 *For an overcomplete subspace dictionary* $\mathfrak{D} = \{\mathfrak{s}_{1:M}\}$ *with* $\mathrm{Span}(\mathfrak{D}) = \mathbb{C}^N$,

$$\begin{aligned} H(\mathfrak{D}) &= \mathbb{C}^N \\ \mathbb{C}^N \backslash H(\mathfrak{D}) &= \varnothing \end{aligned}$$

For a linear independent subspace set $\{\mathfrak{s}_\Lambda\}$ *with* $\mathrm{Span}(\hat{\mathfrak{s}}_\Lambda) = \mathbb{C}^N$,

$$\mathbb{C}^N \backslash H(\mathfrak{s}_\Lambda) \neq \varnothing$$

$H(\mathfrak{s}_\Lambda)$ *is a subspace of* $\mathbb{C}^N$ *and thus a null set. Since* $H(\mathfrak{D})$ *is a union of finite number of null sets,* $H(\mathfrak{D})$ *is also a null set.*

Definition 21 *Vector-Subspace linear independency*

Let $\{\mathfrak{s}_{\Lambda(1)}, \cdots, \mathfrak{s}_{\Lambda(L)}\}$ *be a set of equal-dimensional subspaces. Let the intersection region* $H(\mathfrak{s}_\Lambda)$ *defined as (6.11).*

The set $\{\mathfrak{s}_{\Lambda(1)}, \cdots, \mathfrak{s}_{\Lambda(L)}\}$ *is said to be* vector-subspace-linear-independent*, if for* $\forall i = 1:L$, $\forall p \in \mathrm{Span}(\mathfrak{s}_i)$ *and* $p \notin H(\mathfrak{s}_\Lambda)$*:*

$$\forall \xi_{j \neq i} \in \mathbb{C}^C : \ \textstyle\sum_{j \neq i} \hat{\mathfrak{s}}_j \cdot \xi_j \neq p$$

Lemma 15 *The subspace linear independency (Definition 15) implies vector-subspace linear independency (Definition 21).*

Proof

i. If $\{\mathfrak{s}_{1:L}\}$ *is subspace linear independent, then*
$\forall i = 1 : L$, $\{p \in \mathfrak{s}_i | p \notin H(\mathfrak{s}_{1:L})\} \neq \varnothing$.

ii. follows from definition 21:
If $p \in \mathfrak{s}_i$ *could be represented by the other subspaces, then* $p \in \mathrm{Span}(\mathfrak{s}_{j\neq i})$.
It's contradictory to the two other properties:

$$p \in \mathfrak{s}_i$$
$$p \notin (\mathfrak{s}_i \cap \mathrm{Span}(\mathfrak{s}_{j\neq i})) \subset H(\mathfrak{s}_{1:L}).$$

□

Lemma 16 *Uniqueness and subspace linear independency*
Given subspace dictionary $\mathfrak{D}$ *with subspace Spark* $\mathrm{SSpark}(\mathfrak{D}) = L$.
If a vector $p \in \mathbb{C}^N \backslash \hat{H}(\mathfrak{D})$ *can be represented by both subsets of subspaces* Λ_1 *and* Λ_2, *i.e.*

$$\begin{aligned} p &\in \mathrm{Span}(\hat{\mathfrak{s}}_{\Lambda_1}) \\ p &\in \mathrm{Span}(\hat{\mathfrak{s}}_{\Lambda_2}), \end{aligned}$$

then $\{\mathfrak{s}_{\Lambda_1\cup\Lambda_2}\}$ *must be subspace linear dependent and*

$$|\Lambda_1 \cup \Lambda_2| \geqslant L$$

Proof
We devide $\{\mathfrak{s}_{\Lambda_1}\}$ into one member subspace and the rest: $\{\mathfrak{s}_1\} + \{\mathfrak{s}_{\Lambda_1'}\}$.
We notate $p_1 := \hat{\mathcal{P}}_{\mathfrak{s}_1}(p)$, $p' := \hat{\mathcal{P}}_{\mathrm{Span}(\mathfrak{s}_{\Lambda_1'})}(p) \Rightarrow\ p = p_1 + p'$
If p can also be represented by another subset of subspaces Λ_2 *and* $\{\mathfrak{s}_{\Lambda_1\cup\Lambda_2}\}$ is subspace linear independent, then
$p \in \mathrm{Span}(\hat{\mathfrak{s}}_{\Lambda_2})$,
thus $p' \in \mathrm{Span}(\hat{\mathfrak{s}}_{\Lambda_2})$,
and thus $p_1 = (p - p') \in \mathrm{Span}(\hat{\mathfrak{s}}_{\Lambda_1'\cup\Lambda_2})$.
On the other hand $p_1 \in \mathfrak{s}_1$. Together with $p \notin \hat{H}(\mathfrak{D})$ and lemma 15, it's contradictory to the assumption that $\{\mathfrak{s}_{\Lambda_1\cup\Lambda_2}\}$ is supspace linear indepent.
Hence $\{\mathfrak{s}_{\Lambda_1\cup\Lambda_2}\}$ is must be subspace linear dependent and $|\Lambda_1 \cup \Lambda_2| \geqslant L$.
□

Uniqueness theorem and its proof

Theorem 4 *Uniqueness Condition*

Let $\boldsymbol{p} < \mathbb{C}^N$ be a target vector. Let $\mathfrak{D}$ be a subspace dictionary with the coherence μ^C.

If p can be represented by a set of L subspaces $\{\mathfrak{s}_{\Lambda(1)}, \cdots, \mathfrak{s}_{\Lambda(L)}\}$ with

$$L < \frac{1}{2} \cdot \left(\frac{\sqrt{C}}{\mu^{(C)}} + 1 \right) \tag{6.12}$$

Further if

$$p \notin \hat{H}(\mathfrak{D}), \tag{6.13}$$

then this representation is unique.

Since $\hat{H}(\mathfrak{D})$ is a null-set in $\mathbb{C}^N$, one can also formulate in a loose way:

If (6.12) is true, then it's very likely the unique representation. "Very likely" in the sense of (6.13) and $\hat{H}(\mathfrak{D})$ is a null set in $\mathbb{C}^N$.

Proof

Outline: The proof goes in two stages: (i)We express the uniqueness condition with the quantity of the subspace dictionary spark ($\mathrm{SSpark}(\mathfrak{D})$). (ii)We estimate $\mathrm{SSpark}(\mathfrak{D})$ by the more calculable quantity of subspace mutual coherence.

(i) If one has another subspace set $\{\mathfrak{s}_{\Lambda'}\}$ with $|\Lambda'| \leqslant |\Lambda|$, then according to *Lemma-16* $\{\mathfrak{s}_{\Lambda \cup \Lambda'}\}$ must be linear dependent in the subspace sense. This means

$$2 \cdot |\Lambda|_0 \geqslant |\Lambda \cup \Lambda'|_0 \geqslant \mathrm{SSpark}(\mathfrak{D}) \tag{6.14}$$

where $|\cdot|$ notates the length of the index vector, which is the number of the participating subspaces.

Hence, if p can be represented by a subspace set with the number of members L less than half of the subspace spark, i.e.

$$L < \frac{1}{2} \cdot \mathrm{SSpark}(\mathfrak{D}), \tag{6.15}$$

then the *uniqueness* of this representation is guaranteed, because then (6.14) cannot be hold.

(ii) From *Lemma-12* we get the estimation

$$\mathrm{SSpark}(\mathfrak{D}) > \frac{\sqrt{C}}{\mu^{(C)}} + 1$$

together with (6.15), we get the uniqueness condition in μ^C:

$$L < \frac{1}{2} \cdot \left(\frac{\sqrt{C}}{\mu^{(C)}} + 1 \right)$$

□

6.3.3 Optimality condition theorem

Theorem 5 *Optimality of Subspace-OMP*

We assume the target vector $\boldsymbol{p} \in \mathbb{C}^N$ *has a unique sparsest subspace representation out of the subspace dictionary* $\mathfrak{D} := \{\mathfrak{s}_{\Lambda(1)}, \cdots, \mathfrak{s}_{\Lambda(M)}\}$.

If the number of the participating subspaces m *satisfies the upper bound condition:*

$$m < \frac{1}{2}(\frac{1}{\mu^{(C)}} + 1),$$

then the subspace-OMP algorithm is guaranteed to find the optimal sparsest representation.

After all the preparationary work, the proof of the subspace-OMP optimality theorem follows an analog way as the proof of OMP optimality, (§5.3.3).

The proof outline:

i. We show that a sufficient condition for an exact recovery of the sparsest solution using Subspace-OMP is $\|\hat{\mathfrak{D}}^*_{\text{opt}} \cdot r\|_{2|\infty} > \|\hat{\mathfrak{D}}^*_{\text{rest}} \cdot r\|_{2|\infty}$, where $\hat{\mathfrak{D}}_{\text{opt}}$ is the subspace subdictionary matrix form of the optimal subspace atoms of the sparsest representation and $\mathfrak{D}_{\text{rest}}$ is the one of the non-optimal subspace atoms. $r \in \mathbb{C}^N$ is the residual vector.

ii. We show that $\max_{\omega \notin \text{opt}} \|\hat{\mathfrak{D}}^*_{\text{opt}} \cdot \mathfrak{s}_\omega\|_{2|1} < 1$ implies $\|\hat{\mathfrak{D}}^*_{\text{opt}} \cdot r\|_{2|\infty} < ||\hat{\mathfrak{D}}^*_{\text{rest}} \cdot r\|_{2|\infty}$

iii. We derive an upper bound of $\max_{\omega \notin \text{opt}} \|\hat{\mathfrak{D}}^*_{\text{opt}} \cdot \mathfrak{s}_\omega\|_{2|1}$ using the subspace cumulative coherence: $\max_{\omega \notin \text{opt}} \|\hat{\mathfrak{D}}^*_{\text{opt}} \cdot \mathfrak{s}_\omega\|_{2|1} < 1 - \frac{1-\mu_1^{(C)}(m-1)-\mu_1^{(C)}(m)}{1-\mu_1^{(C)}(m-1)}$

iv. We combine i. ii. iii. to get $\mu_1^C(m-1) + \mu_1^C(m) < 1$ as an optimality condition. Since $\mu_1^C(m) \leqslant m \cdot \mu^C$, we have the theorem statement $m < \frac{1}{2}(\frac{1}{\mu^C} + 1)$.

The complete proof is based on three lemmata covering accordingly the steps i, ii, iii above.

Lemma 17 *Let* $\hat{\mathfrak{D}}_{\text{opt}}$ *be the matrix form of the optimal subspace atoms used by the sparsest representation of* p, $\hat{\mathfrak{D}}_{\text{rest}}$ *be the matrix form of the non-optimal subspace atoms,* r *be the residual vector calculated in the subspace-OMP-algorithm. Then, the condition* $\|\mathfrak{D}^*_{\text{rest}} r\|_\infty < \|\mathfrak{D}^*_{\text{opt}} r\|_\infty$ *is a sufficient condition for the Subspace-OMP method to always find the sparsest representation.*

Proof

In the subspace OMP algorithm, (*Algorithm 5*), the initial value of the residual are set to be $r := p \in \mathrm{Span}(\hat{\mathfrak{D}}_{\mathrm{opt}})$. The optimal atom of each iteration step is chosen by the maximal inner-product of the subspace atoms and the residual vector, i.e. the $\|\cdot\|_{2|\infty}$-norm.

$$\|\mathfrak{D}^*_{\mathrm{rest}} r_n\|_{2|\infty} < \|\mathfrak{D}^*_{\mathrm{opt}} r_n\|_{2|\infty}.$$

It guarentees that from n-th iteration to $(n+1)$-th iteration, only atoms from the optimal set will be chosen.

Since $\boldsymbol{p} \in \mathrm{Span}(\mathfrak{D}_{\mathrm{opt}})$ and all the chosen atoms are from the optimal set, the resulting new residual vector also lies in the span of the optimal atoms: $r_{n+1} \in \mathrm{Span}(\mathfrak{D}_{\mathrm{opt}})$.

After M steps, all elements of the optimal set are chosen. The aimed optimal representation is then obtained by the least squares step.

□

Lemma 18 $\forall r \in \mathrm{Span}(\hat{\mathfrak{D}}_{\mathrm{opt}})$*, a sufficient condition for*

$$\|\hat{\mathfrak{D}}^*_{\mathrm{rest}} r\|_{2|\infty} < \|\hat{\mathfrak{D}}^*_{\mathrm{opt}} r\|_{2|\infty} \tag{6.16}$$

is

$$\max_{\omega \notin \mathrm{opt}} \|\hat{\mathfrak{D}}^*_{\mathrm{opt}} \mathfrak{s}_\omega\|_{2|1} < 1$$

Proof

Since $r \in \mathrm{Span}(\hat{\mathfrak{D}}_{\mathrm{opt}})$, its projection on $\mathrm{Span}(\hat{\mathfrak{D}}_{\mathrm{opt}})$ is r itself. $\hat{\mathcal{P}}_{\mathrm{opt}}(r) = r$, where $\hat{\mathcal{P}}_{\mathrm{opt}}$ is the projection operator onto $\mathrm{Span}(\hat{\mathfrak{D}}_{\mathrm{opt}})$. $\hat{\mathcal{P}}_{\mathrm{opt}} = \hat{\mathfrak{D}}_{\mathrm{opt}} \cdot \hat{\mathfrak{D}}^\dagger_{\mathrm{opt}}$. Since $\hat{\mathcal{P}}_{\mathrm{opt}}$ is conjugate symmetric: $\hat{\mathcal{P}}_{\mathrm{opt}} = (\hat{\mathfrak{D}}^\dagger_{\mathrm{opt}})^* \cdot \hat{\mathfrak{D}}^*_{\mathrm{opt}}$. We get:

$$\|\hat{\mathfrak{D}}^*_{\mathrm{rest}} r\|_{2|\infty} = \|\hat{\mathfrak{D}}^*_{\mathrm{rest}} \cdot \hat{\mathcal{P}}_{\mathrm{opt}} \cdot r\|_{2|\infty} = \|\hat{\mathfrak{D}}^*_{\mathrm{rest}} \cdot (\hat{\mathfrak{D}}^\dagger_{\mathrm{opt}})^* \cdot \hat{\mathfrak{D}}^*_{\mathrm{opt}} \cdot r\|_{2|\infty}$$

Together with the triangle inequation of $\|\cdot\|_{2|\infty}$, we get

$$\|\hat{\mathfrak{D}}^*_{\mathrm{rest}} r\|_{2|\infty} < \|\hat{\mathfrak{D}}^*_{\mathrm{rest}} \cdot (\hat{\mathfrak{D}}^\dagger_{\mathrm{opt}})^*\|_{2|\infty} \cdot \|\hat{\mathfrak{D}}^*_{\mathrm{opt}} \cdot r\|_{2|\infty} \tag{6.17}$$

Then the sufficient condition for (6.16) is

$$\|\hat{\mathfrak{D}}^*_{\mathrm{rest}} \cdot (\hat{\mathfrak{D}}^\dagger)^*\|_{2|\infty} < 1 \tag{6.18}$$

We set the connection between the $\|\cdot\|_{2|\infty}$- and $\|\cdot\|_{2|1}$-norm. Since the ∞-norm of a matrix is the maximum of the absolute row-sum: $\|A\|_\infty = \max_i \left(\sum_j |a_{ij}|\right)$, we get the following relation:

$$\|\hat{\mathfrak{D}}^*_{\mathrm{rest}} (\hat{\mathfrak{D}}^\dagger_{\mathrm{opt}})^*\|_{2|\infty} = \left\| \left(\hat{\mathfrak{D}}^\dagger_{\mathrm{opt}} \cdot \hat{\mathfrak{D}}_{\mathrm{rest}} \right)^* \right\|_{2|\infty} = \max_{\omega \notin \mathrm{opt}} \|\hat{\mathfrak{D}}^*_{\mathrm{opt}} \hat{\mathfrak{s}}_\omega\|_{2|1} \tag{6.19}$$

Combining (6.17) and (6.19), the neccessary condition for (6.16) reads

$$\max_{\omega \notin \text{opt}} \|\hat{\mathfrak{D}}_{\text{opt}}^{\dagger} \hat{\mathfrak{s}}_{\omega}\|_{2|1} < 1 \tag{6.20}$$

□

So we've derived the sufficient condition for (6.16) as a up boundary for $\max_{\omega \notin \text{opt}} \|\hat{\mathfrak{D}}_{\text{opt}}^{\dagger} \mathfrak{s}_{\omega}\|_{2|1}$. Now we estimate the quantity $\max_{\omega \notin \text{opt}} \|\hat{\mathfrak{D}}_{\text{opt}}^{\dagger} \mathfrak{s}_{\omega}\|_{2|1}$ itself by the subspace dictionary cumulative coherence μ_1^C.

Lemma 19 *Let the number of the optimal subspace atoms be M, then*

$$\max_{\omega \notin \text{opt}} \|\hat{\mathfrak{D}}_{\text{opt}}^{\dagger} \mathfrak{s}_{\omega}\|_{2|1} < 1 - \frac{1 - \mu_1^{(C)}(m-1) - \mu_1^{(C)}(m)}{1 - \mu_1^{(C)}(m-1)}. \tag{6.21}$$

Proof

a) We express the quantity by the Gram matrix, and apply the triangle inequalty:

$$\begin{aligned} \max_{\omega \notin \text{opt}} \|\hat{\mathfrak{D}}_{\text{opt}}^{\dagger} \mathfrak{s}_{\omega}\|_{2|1} &= \max_{\omega \notin \text{opt}} \|(\hat{\mathfrak{D}}_{\text{opt}}^{*} \cdot \hat{\mathfrak{D}}_{\text{opt}})^{-1} \cdot \hat{\mathfrak{D}}_{\text{opt}}^{*} \cdot \hat{\mathfrak{s}}_{\omega}\|_{2|1} \\ &\leqslant \|(\hat{\mathfrak{D}}_{\text{opt}}^{*} \cdot \hat{\mathfrak{D}}_{\text{opt}})^{-1}\|_{2|1} \cdot \max_{\omega \notin \text{opt}} \|\hat{\mathfrak{D}}_{\text{opt}}^{*} \cdot \hat{\mathfrak{s}}_{\omega}\|_{2|1} \end{aligned}$$

b) Using the boundary estimation for the inverse Gram matrix G_{opt} (Lemma-10), we get

$$\|(\hat{\mathfrak{D}}_{\text{opt}}^{*} \cdot \hat{\mathfrak{D}}_{\text{opt}})^{-1}\|_{2|1} = \|G_{\text{opt}}^{-1}\|_{2|1} < \frac{1}{1 - \mu_1^{(C)}(m-1)}$$

c) Due to the definition of cumulative coherence $\mu_1^{(C)}(m)$, we get

$$\max_{\omega \notin \text{opt}} \|\hat{\mathfrak{D}}_{\text{opt}}^{*} \cdot \hat{\mathfrak{s}}_{\omega}\|_{1} = \max_{\omega \notin \text{opt}} \left(\sum_{\lambda \in \text{opt}} \|\hat{\mathfrak{s}}_{\omega}^{*} \cdot \hat{\mathfrak{s}}_{\lambda}\|_2 \right) \leqslant \mu_1^{(C)}(m)$$

We combine a), b), c) to get (6.21).

□

Proof *of Theorem 5*

Combining the sufficient condition (6.18) for (6.16) with the upper bound estimation (6.21), we get the sufficient condition for (6.16):

$$\frac{1 - \mu_1^{(C)}(m-1) - \mu_1^{(C)}(m)}{1 - \mu_1^{(C)}(m-1)} < 0. \tag{6.22}$$

A sufficient condition for (6.22) in turn is

$$\mu_1^{(C)}(m-1) + \mu_1^{(C)}(m) < 1 \tag{6.23}$$

Together with *lemma-17*, we get that (6.23) is a sufficient condition for the subspace-OMP to find the optimal solution.

The last step is to replace the subspace cumulative coherence with the more calculatable mutual coherence $\mu^{(C)}$. Since $\mu_1^{(C)}(m-1) < (m-1) \cdot \mu^{(C)}$ (*lemma-9*), we get the optimality condition for subspace-OMP:

$$m < \frac{1}{2}(\frac{1}{\mu^{(C)}} + 1)$$

□

Part III

The Transfer from Math. to MRI

Chapter 7

Application Setting 1: Sparse Single Channel RF Transmit

7.1 Recall the Problem: Adaptive Single Channel Transmit

In the Chapter 4 we discussed the Spatial Selective Excitation Problem $\boldsymbol{P}_{\mathrm{SSEP}}$. The task is to design two waveforms, $\mathrm{RF}(t)$ and $g(t)$ to excite a spatially distributed spin ensemble from their ground state to an user given spatial magnetization target profile.

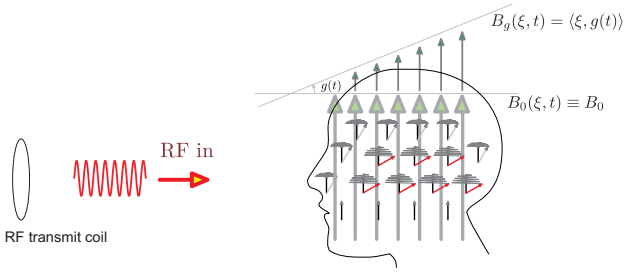


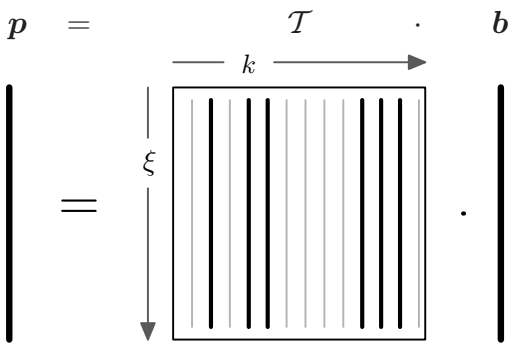
Figure 7.1: Schematic for selective excitation process

Else than in the imaging problem, where we don't have concrete knowledge of the images to be acquired, in the excitation problem we have the full knowledge of the target magnetization profile to be excited. This full prior information of the target profile leads us to the adaptive concept. We want to individually shorten the excitation process tailored to the individual target profile.

In the STA model the conventional non-adaptive paradigm reduces $\boldsymbol{P}_{\text{SSEP}}$ to two subproblems $\boldsymbol{P}_{\text{kTrav}}(\boldsymbol{\kappa}_{\text{Nyq}})$ and $\boldsymbol{P}_{\text{RF-design}}$. (See section 4.1.) The $\boldsymbol{P}_{\text{kTrav}}(\boldsymbol{\kappa}_{\text{Nyq}})$ is a task of fast covering the Nyquist sampling grid with the traveling velocity and acceleration constraints. The $\boldsymbol{P}_{\text{RF-design}}$ is a linear representation problem using the according Fourier vectors.

Also under the STA model, the proposed adaptive paradigm approaches the $\boldsymbol{P}_{\text{SSEP}}$ with three subproblems $\boldsymbol{P}_{\text{sparse}}$ (4.25), $\boldsymbol{P}_{\text{kTrav}}$ (4.15) and $\boldsymbol{P}_{\text{RF-design}}$ (**??**). The $\boldsymbol{P}_{\text{sparse}}$ is a sparse approximation problem using a pure Fourier dictionary. Its task is to represent the target profile vector $\boldsymbol{p}$ with as less Fourier harmonics as possible.

The $\boldsymbol{P}_{\text{sparse}}$ can be schematically demonstrated as following:



After selecting out the most important k sampling locations, we translate the sampling location $\{k_{1:L}\}$ to feasible physical control of the gradient waveform $g(t)$, with the gradient and slew-rate boundary: $g(t) \leqslant G_{\max}$ and $\dot{g}(t) \leqslant S_{\max}$. Since $k(t) = \int_t^{T_f} g(\tau)d\tau$, this step is equivalent to a problem of fast covering the obligatory k space locations with the traveling velocity and the acceleration constraints. This is the task of subproblem $\boldsymbol{P}_{\text{kTrav}}$.

Compared with the problem $\boldsymbol{P}_{\text{kTrav}}(\boldsymbol{\kappa}_{\text{Nyq}})$, the $\boldsymbol{P}_{\text{kTrav}}$ doesn't assume regularly distributed obligatory locations like the Nyquist grid. Usually the input obligatory locations are quite irregularly distributed. Thus solving $\boldsymbol{P}_{\text{kTrav}}$ is not easy.

In section 7.2 we'll discuss the question of choosing the appropriate dictionary for $\boldsymbol{P}_{\text{sparse}}$: Concretely, whether the overcomplete Fourier dictionary or the orthonormal Fourier dictionary. We will come to the conclusion of choosing the orthonormal Fourier dictionary.

In section 7.3 we'll discuss the numerical algorithm of $\boldsymbol{P}_{\text{sparse}}$ tailored to

the orthogonal Fourier dictionary. We will come to an algorithm with very intuitive physical interpretation of the *adaptive energy threshold*.

In section 7.4 we'll discuss our approach for the $\boldsymbol{P}_{\text{kTrav}}$. We follow two alternative empirical methods, one from convex optimization perspective, one from time optimal control perspective.

In the section 7.5 we stablish the validation strategy of three stages.

In section 7.6 we'll validate the undersampling-enhancing property of the adaptive energy threshold algorithm approaching $\boldsymbol{P}_{\text{sparse}}$.

In section 7.7 we'll validate the actual acceleration effect of the whole concept via simulation of the relaxation free Bloch equation.

In section 7.8 we'll validate the praxis applicability via the "real-world" phantom experiment on a GE 3T MRI scanner.

In section 7.9 we'll demonstrate the adaptive capability of the proposed approach via several examples.

7.2 Dictionary Choosing: ON vs. Overcomplete Fourier Dictionary

The choose of dictionary for $\boldsymbol{P}_{\text{sparse}}$ is physically restricted to the pure Fourier dictionary. Hence we indeed choose between the overcomplete Fourier dictionary or complete orthonormal Fourier basis.

We consider this question from a pragmatic benefit-cost perspective. We ask if the additional complexity of the overcomplete Fourier dictionary can be justified with the benefit of additional acceleration effect.

We at first compare the two dictionaries from a theoretical point of view. Then we compare them at a heuristic point of view with numerical example.

7.2.1 Theoretical guarantee of finding the optimal sparse solution

In chapter 5 we discussed the theoretical result concerning the optimality condition for the two major heuristic methods for sparse approximation. After the work of J. Tropp et al.[Tro04b] and D. Donoho et al.[DETR06] both OMP method and L1-minimization methods are guaranteed to find the sparsest solution when the target vector has indeed a sparse representation:

$$\|\boldsymbol{c}\|_0 = (\mu^{-1} + 1)/2$$

The dictionary coherence property μ is essential here.

If the overcomplete dictionaries are constructed by concatenating multiple orthogonal dictionaries, the dictionary coherence μ could still be well bounded, $\mu \geqslant \sqrt{\frac{1}{N}}$ (Proposition 1). However, the dictionary coherence μ of the pure Fourier dictionary grows unfortunately very fast as the overcompleteness increases.

To demonstrate that we take an example of an overcomplete Fourier pure dictionary with the frequency sampling rate twice than required by the Nyquist theorem. We have atoms:

$$\varphi_j(\xi) = \exp\{\frac{2\pi}{\nu \cdot N} \cdot i \cdot \xi\}, \tag{7.1}$$

where ξ is the spatial coordinate, N is the dimension of the target vector, ν is the oversampling rate regarding the Nyquist sampling grid in the frequency domain. The factor $\nu = 1$ is equivalent to an ON Fourier basis fulfilling Shannon-Nyquist theorem. $\nu > 1$ is the oversampling factor. For $\nu = 2$ one has a twice oversampled dictionary. This twice overcomplete dictionary has a coherence $\mu(N) = \frac{1}{N \cdot \sin(\pi/N)}$ [DETR06]. For the limit $\frac{\pi}{N} \ll 1$ the μ converges to $1/\pi$:

$$\lim_{N \gg 1}(\mu) = \frac{1}{\pi} = 0.3183$$

For example for a 2D profile $N = 33 \times 33 = 1089$, $\mu(1089) \doteq 0.3183$.

The according critical sparsity for a theoretical guarantee then turns to be

$$\|c\|_0 \leqslant \frac{1}{2} \cdot (\frac{1}{\mu(1089)} + 1) \approx 2.07.$$

This means the theory can only offer an optimality guarantee when the 2D profile can be represented by maximally two atoms. This is practically never the case.

Conclusion: for a overcomplete pure Fourier dictionary, the current theoretical result cannot provide praxis meaningful help.

Remark: The theorems of the theoretical guarantee play indeed the worst scenario. It doesn't mean for the target vectors beyond that critical sparsity the two numerical methods are unusable, but just there is no guarantee any more for optimality. Also not be said is, when fails, how far away it is from the optimal solution, whether it is still somehow acceptable or disastrous.

On the other hand, Orthonormal basis always has optimality guarantee, even when it is not sparse, (Corollary 1). due to $\mu(\mathrm{ON}) = 0$, the critical sparsity level turns to be infinity:

$$\|\boldsymbol{c}\|_0 \leqslant \frac{1}{2} \cdot (\frac{1}{0} + 1) = \infty$$

So the optimality is guaranteed for any target vector.

Remark: The overcomplete dictionaries are weak at the guarantee of optimum-finding, but it has generally better optimum. In contrast, orthogonal basis have generally less good optimum, while it always has the guarantee of being able to find the optimum.

7.2.2 Benefit vs. cost — heuristic point of view

We compare the overcomplete and orthonormal dictionaries in a heuristic point of view.

We recall that the actual acceleration is achieved by two steps: 1) sparsifying the obligatory sampling locations from the Nyquist grid, 2) designing a traveling trajectory in k space covering those sparse obligatory locations. Therefore the solution quality offered by the dictionary should be checked accordingly in two levels: 1. sparsity of the k space obligatory locations, 2. the actual traveling duration of the k space trajectory, which is indeed the excitation duration we want to shorten.

We take the heart shape as our target magnetization profile as in (fig-1.4-top-left). We choose the user tolerance 10%. We perform the tests for the two candidate dictionaries: *Dictionary-1:* the twice oversampled Fourier dictionary, $\nu = 2$ in (7.1); *Dictionary-2:* the orthonormal Fourier basis. For the sparse approximation we use the OMP algorithm.

Dictionary-1: **Overcomplete Fourier Basis,** $\nu = 2$ (fig-7.2-top-row)

> For the sparse approximation level, the OMP calculation using the orthonormal Fourier basis ends up with **56** obligatory k space location. The computational time for that is *176* s.
>
> For the actual acceleration level, the k trajectory design ends up with **1070** sampling points with $4\mu s$ sampling distance. This k trajectory has accordingly a traveling duration of *4.2ms*. (For more detail regarding k space traveling design step, please see §-7.4.)

Dictionary-2: **Orthogonal Fourier basis** (fig-7.2-bottom-row)

> For the comparison level-1, the sparse approximation, the OMP calculation using the complete orthogonal Fourier dictionary ends up with **60** atoms. The computational time for selecting out these 60 locations is *60*s.
>
> For the comparison level-2, the actual acceleration, the k trajectory design covering these 60 k-locations ends up with **458** sampling points

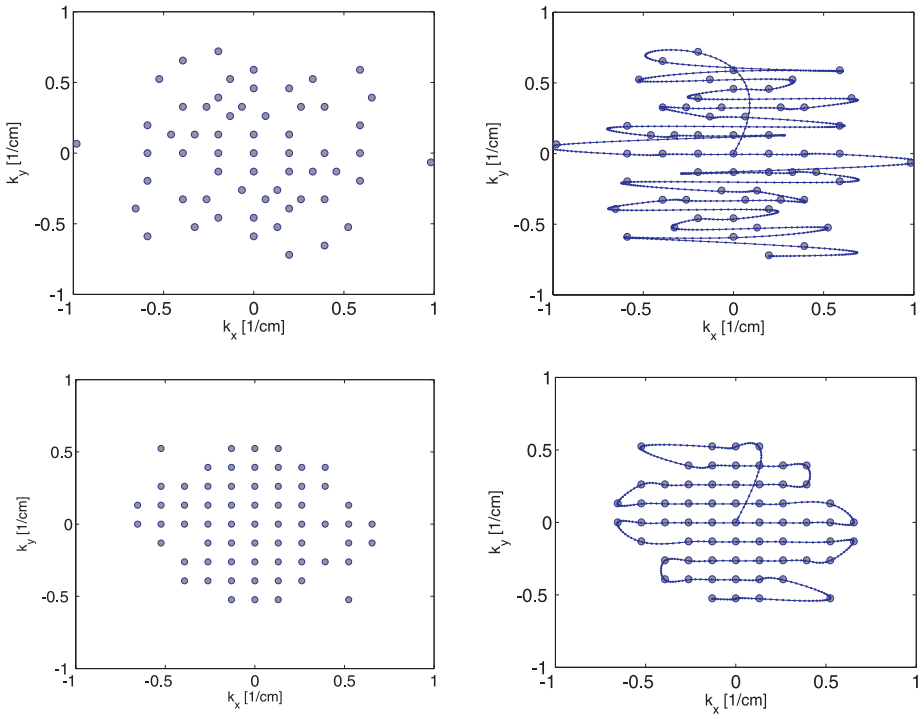


Figure 7.2: Overcomplete- vs. ON-dictionary: k space sampling (*Top-left:* Sparse approximation result with overcomplete dict. *Top-right:* designed trajectory covering the overcomplete sparse approx. results. *Bottom-left:* sparse approx. result with ON dict. *Bottom-right:* designed trajectory covering the ON sparse approx. results)

with $4\mu s$ sampling distance. The k trajectory has accordingly a traveling duration of *1.83ms*.

Specially for the ON Fourier basis a tailored algorithm can further reduce the numerical cost. (For details please see §-7.3.) As the results one get the same output of k-locations and k-trajectory as above. However the computational time is much shorter: *3*s

To summary up we bring together the facts into table 7.1.

If we only look at the sparsity of the k-sampling-locations, the overcomplete dictionary gives slightly sparser solution, 56 vs. 60.

	Overcomplete	ON
# k locations:	56	60
# k trajectory:	1070 (=4.2ms)	458 (=1.83ms)
computational time (OMP):	176 s	60 s
specified for ON:		3s

Table 7.1: The numerical cost comparison

If we look at the more important quantity, the traveling duration of the k-trajectory, the solution of the overcomplete dictionary has surprisingly even longer duration than the ON-dictionary solution with slightly more obligatory locations to cover: 4.2ms vs. 1.83ms.

There are basically two reasons that to recover 56 locations costed more time than to cover 60 locations:

1. The 56 locations from the overcomplete dictionary are more loosely distributed than the second set of 60 locations.

2. The 56 locations are more irregular distributed than the 60 locations from the orthogonal Fourier basis. Solving $\boldsymbol{P}_{\mathrm{kTrav}}$ is more demanding with more irregular distributed locations.

Remark: The big gap between the durations (458 vs. 1070) might be reducible to certain level if one has more advanced k traveling design strategy. However the core massage will still remain: *the slight advantage in the sparsity, like 56 vs. 60, doesn't guarantee the actual acceleration at all.*

Let's take a look a the computational cost.

The numerical cost of OMP is of $\mathcal{O}(L \cdot N \cdot M)$. Here N is the dimension of the target vector, M is the number of the atoms in the dictionary, L is the number of iteration.

If one use OMP algorithm, the difference of the numerical cost comes basically from the number of atoms in the dictionary: M. For our overcomplete Fourier dictionary case we twice oversampled in both k_x and k_y direction, thus it has 4×the number of the atoms as the ON Fourier basis elements. Therefore we also expect a numerical cost difference up to this fact. The numerical test result roughly confirmed this estimation (176s vs 60s, table-7.1).

For the orthogonal dictionary, the OMP algorithm can be further optimized to a much cheaper algorithm based on orthogonal decomposition. Using this tailored algorithm for ON dictionary one can reduce the numerical cost further from $\mathcal{O}(L{\cdot}N{\cdot}M)$ to $\mathcal{O}(N^2)$, with FFT even to $\mathcal{O}(N{\cdot}\ln(N){+}L{\cdot}N)$. In the example above it reduced the computational time from 60s to 3s. We will discuss this ON basis tailored algorithm in more detail in next section.

7.2.3 Conclusion

Combine all these factors, especially the cost benefit comparison, it turns out clearly that introducing the overcomplete Fourier dictionary will bring large additional complexity and numerical burden, while the benefit of it is very small or even not guaranteed. It can well happen that regarding the actual acceleration the overcomplete dictionary delivers even less optimal solution than the ON dictionary.

So we will chose the complete orthonormal Nyquist Fourier basis as our dictionary for the application setting of single channel RF transmit.

7.3 The Algorithm Tailored for Orthogonal Dictionary

For the special case of orthogonal dictionary the OMP algorithm can not only guarantee the optimality, but also be further optimized to reduce the cost from $\mathcal{O}(L \cdot N^2)$ to $\mathcal{O}(N^2)$. In this section we will introduce this ON dictionary tailored algorithm.

7.3.1 The algorithm of adaptive energy threshold

For orthogonal dictionary, the greedy algorithm OMP can be reduced to an orthogonal decomposition step and an adaptive break check step. It results

the *Algorithm 7.3* with the following input and output:

$$(\boldsymbol{Input}:)\quad \begin{cases} \boldsymbol{p} & \text{(the normalized target profile)} \\ \Phi = [\varphi_1, \ldots, \varphi_N] & \text{(the orthonormal dictionary)} \\ 0 \leqslant \mathrm{TOL} < 1 & \text{(user tolerance)} \end{cases}$$

$$(\boldsymbol{Output}:)\quad \begin{cases} \Lambda & \text{(the support index)} \\ \boldsymbol{b} \in \mathbb{C}^{\Lambda} & \text{(the dictionary)} \\ \textbf{Required} & \|\boldsymbol{p} - \Phi_{\Lambda} \cdot \boldsymbol{b}\| \leqslant \mathrm{TOL} \end{cases}$$

Algorithm 4 Adaptive Energy Threshold (AET)

1: {Orthogonal decomposition}

$$\boldsymbol{b} \leftarrow \Phi^* \cdot \boldsymbol{p}$$

2: Order the index vector Λ after $|b_{\Lambda(1)}| > \ldots > |b_{\Lambda(N)}|$;

3: {Init. for adaptive Break Check}

$$l \leftarrow 1, \mathrm{res} \leftarrow \boldsymbol{p}$$

4: {Update the residual}

$$\mathrm{res} \leftarrow \mathrm{res} - \varphi_{\hat{\Lambda}(l)} \cdot b_{\hat{\Lambda}(l)}$$

5: **if** $\| \mathrm{res} \|_2 > \mathrm{TOL}$ **then**
6: $l \leftarrow l + 1$; $\rightarrow$ Step 3
7: **else**
8: **return** $\Lambda \leftarrow \{\Lambda(1), \cdots \Lambda(l)\}$, $\boldsymbol{b} \leftarrow \{b_{\Lambda(1)}, \cdots b_{\Lambda(l)}\}$ with success!
9: **end if**

7.3.2 The physical interpretation

This algorithm has a very intuitive physical interpretation: *Adaptive RF energy threshold*.

The physical meaning of the orthogonal decomposition coefficient $\boldsymbol{b} = (b_1, \ldots, b_N)$ is the complex RF waveform value as a function of the k sampling locations. The quantity b_j^2 has the meaning of the energy of the RF waveform at the j-th k sampling point. Therefore ranking the index Λ after $|b_{\Lambda(1)}| \geq \ldots \geq |b_{\Lambda(N)}|$, is nothing else than ranking the k locations after their corresponding RF energy. Choosing only the first L sampling-locations in k space with higher RF energy means nothing else than a cut-off after an energy threshold. The number L is determined adaptively by the criterion whether the user accuracy requirement is reached with the current sampling locations in use. It is equivalent to an adaptively decreasing threshold of the RF energy.

Due to this physical interpretation, we call this method the *Adaptive Energy Threshold* method (AET).

7.3.3 The equivalence to the normal OMP

The essential point here is: Because of the orthogonality of the atom-vectors, the projections of the residual vectors on the individual atoms will never change from iteration step to iteration step. Indeed, also due to the orthogonality, the projections of the residual vector on the individual atoms are always identical to the projection of the target profile on the individual atoms.

Therefore it allows us to calculate the projection coefficient regarding each atoms once for all at the very beginning of the algorithm. It is the orthogonal decomposition step. After that we iteratively check the residuals for break and terminate the algorithm by sufficient accuracy.

After the corollary-1 in §-5.3 OMP always finds optimum for orthogonal dictionary. Due to the equivalence to the OMP, the same is true for the *Algorithm 4: AET*.

7.3.4 The numerical cost

The numerical cost of this algorithm consists basically of two parts. i) The orthogonal decomposition, step 1. It has the numerical cost of $\mathcal{O}(N^2)$. ii) The successive residual check, step 4-9. It has the numerical cost of $\mathcal{O}(L \cdot N)$. The L notates the number of the iterations, where $L < N$ due to the convergence nature of the algorithm. Thus the total numerical cost has the order of $\mathcal{O}(N^2)$.

We compare it with the cost of the general OMP algorithm, which is $\mathcal{O}(L \cdot N \cdot M)$. Since we've chosen the complete Fourier basis as our dictionary, we have $M = N$. Thus the cost of OMP algorithm is $\mathcal{O}(L \cdot N^2)$.

For example, for a solution with 60 participating atoms L = 60. Thus the AET algorithm has a cost reduction of about factor 60. This the reason that in the example in previous section the numerical cost dramatically reduced from 60s to 3s by applying the AET algorithm.

This algorithm can be further optimized by

1. using FFT in the orthogonal decomposition step: It reduces the $\mathcal{O}(N^2)$ cost to $\mathcal{O}(N \cdot \ln N)$.

2. applying Parseval's theorem: The residual-L_2-norm in the spatial domain is simply the L_2-norm of the rest-coefficient-vector: $\|(b_{j+1}, \ldots b_N)\|_2$. In this way the cost in part-ii) is then:

$$L \cdot N - (1 + L) \cdot \frac{L}{2} = N + (N - 1) + \ldots + (N - L)$$

However in our operating order of magnitude ($N = \mathcal{O}(10^3)$) these slight improvements in the numerical efficiency doesn't make essential difference any more.

Remark: The Parseval's theorem is only applicable if one uses the L_2 norm as error measure. If one wants to use more flexible error measure, e.g. the magnitude only error: $\|(|b_1|, \ldots, |b_N|)\|_2$, one still needs to calculate the errors in the spatial domain.

7.4 The Gradient Design, k Space Traveling Problem

7.4.1 The $P_{\mathbf{kTrav}}$ problem

We recall that our task is to design the gradient waveform $g(t)$ and the RF waveform. In the STA model we introduced the substitutional quantity $k(t) \in \mathbb{C}$ to replace the gradient waveform: $k(t) = \int g(\tau) d\tau$. This substitution helped us to set the connection with the Fourier transform formulism and transferred the problem to a linear representation problem.

On the one hand, in the k space picture the linear representation formulism using Fourier vector helps us to find certain obligatory k space sampling locations. It thus also delivers the according criterion for the gradient waveform, so that a successful excitation to the target magnetization profile can be guaranteed.

On the other hand, the gradient waveform must subject to the hardware feasible constraints of the maximal gradient value, $g(t) \leqslant G_{\max} < \infty$, and

the maximal slew-rate, $\dot{g}(t) \leqslant S_{\max} < \infty$. These two gradient hardware constraints have the equivalent form in the k space picture that the k space traveling must subject to the velocity constraint, $\dot{\kappa}(t) \leqslant C_1 < \infty$, and the acceleration constraints, $\ddot{\kappa}(t) \leqslant C_2 < \infty$.

Till now we've selected out the sparse obligatory sampling locations adaptive to the target profile and TOL. The next step is to approach the fast covering $\boldsymbol{\kappa}$ problem, $\boldsymbol{P}_{\text{kTrav}}$ (4.15).

If the obligatory locations to be covered, $\boldsymbol{\kappa} := \{\kappa_1, \cdots, \kappa_L\}$, are fully random, the problem $\boldsymbol{P}_{\text{kTrav}}$ can be seen as a special form of the well-known traveling salesman problem with the velocity and acceleration constraints. We all know that the traveling salesman problem is NP hard.

However, our target is not to solve the $\boldsymbol{P}_{\text{kTrav}}$ optimally, but to find a sufficiently good solution somehow, such that the sparsity of the obligatory locations can be sufficiently transferred to the actual acceleration of the k space traveling. The acceleration of the k space traveling duration is exactly the acceleration of the whole excitation process.

We approach the $\boldsymbol{P}_{\text{kTrav}}$ in two steps:

i. Connecting the set of the unordered obligatory locations $\boldsymbol{\kappa}$ to a ordered sequence of locations $\hat{\boldsymbol{\kappa}}$.

ii. Designing $k(t)$ covering the location-sequence $\hat{\boldsymbol{\kappa}}$ and fulfilling the velocity and the acceleration constraints at the same time. The traveling duration should be as short as possible.

7.4.2 Step-i: connect the dots

To order the set of the obligatory locations to a sequence, we following a very simple strategy to connect the locations line by line in the k-space with a alternating line direction. (In MRI terminology it can be called as a EPI-style.)

It's obvious that this strategy doesn't always give the best solution. But since the obligatory locations are selected from a regular grid (Nyquist sampling grid), one has reason to hope that in most cases it will deliver a reasonable sequences for the step-ii) to start with. Empirically, together with the heuristic approach in the step-ii), it delivers a sufficiently good solution to transfer the sparsity to the actual acceleration. In fig 7.3 is an example of how we connect a set of 55 unordered locations to a sequence.

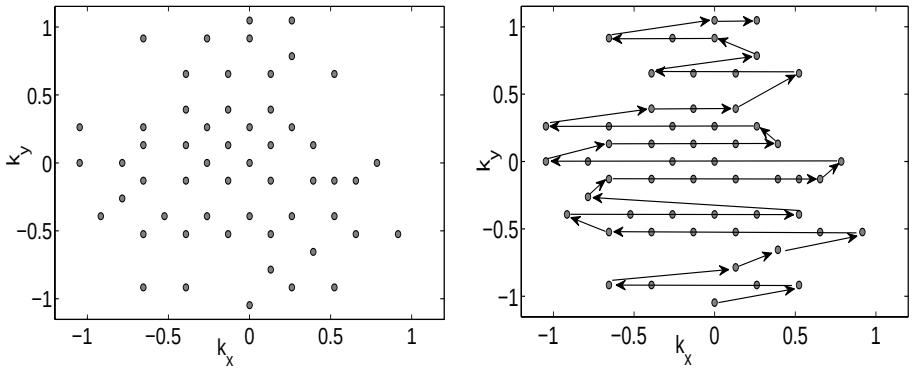


Figure 7.3: Example of connecting the obligatory locations in k space

7.4.3 Step-ii: fast traveling covering the locations-sequence

To fast traveling through the sequence of the obligatory locations subject to the velocity and acceleration constraints, we use two alternative heuristic approaches. The two approaches are based on convex optimization and optimal control respectively. We will discuss them one by one.

Local approach

To introduce our local approach, we starts with a intermediate task of designing a time optimal trajectory connecting only two sampling locations. If the starting and the final velocities are given, the problem can be reduced to a linear programming problem [HNC04]. Since linear programming is a standard convex optimization problem, the optimal solution can be obtained using standard convex optimization software like (CVX, Stanford [GB09]).

Our actual task is to design a trajectory traveling through a sequence of more than two locations. Only the velocities of the initial and final locations are given. All velocities of the interior locations, both magnitude and direction, are to be optimized too. However, if these velocities at the interior sampling locations are given, the rest of the problem is then straightforward. One just solve $(L-1)$ linear programming problems.

Our pragmatic approach is to empirically set the velocity values at the interior locations using some simple calculation rule from the information of the previous and the later k sampling locations. The experience with numerous numerical tests shows that even this kind of very simple strategy can deliver sufficiently good result.

For solving the linear programming problem, we used the nice convex

optimization software package from M. Grant and S. Boyd [GB09].

In fig 7.4 is an example applying the local approach. It use the input location sequence as in the previous example (fig 7.3):

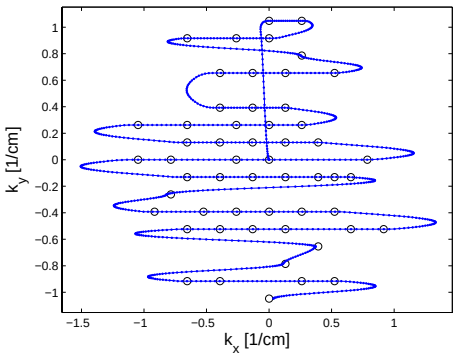


Figure 7.4: Result of $\boldsymbol{P}_{\mathrm{kTrav}}$ using local approach

It ended up with 4.2ms to cover the 55 obligatory locations. (For comparison: to cover the whole Nyquist grid with 1089 obligatory locations, it ended up with 16ms)

Global approach

To introduce our global approach, we start with a given continuous curve in k space, $k_y = C(k_x)$. The time optimal traveling $k(t)$ through C can be calculated using optimal control approach [LKP08]. In this case the time optimality can be guaranteed.

For our actual task $\boldsymbol{P}_{\mathrm{kTrav}}$ we need to at first construct a continuous curve covering the given location sequence ourselves. The concrete shape of the continuous curve covering the discrete location sequence influences the time optimality of the final traveling duration.

A straightforward way to connect a sequence of discrete locations to a continuous curve is the Spline fitting. In our implementation we do take this pragmatic approach of the Spline fitting to at first construct a continuous curve. Then we calculate the traveling movement $k(t)$ through it. The heuristics shows this approach also delivers sufficiently good result to transfer the sparsity to actual acceleration.

For implementation, we basically use the nice matlab package from Miki Lustig, "timeOptimalGradient" [Lus].

However, it needs to be mentioned that the Spline fitting might be a pragmatic choice but in general it is not an optimal strategy to construct the continuous curve. Because on the one hand the Spline fitting corresponds

to finding a continuous curve covering a discrete location sequence with the property of minimal L_2 norm, on the other hand one can easily verify that for fast traveling with the velocity constraints, the optimal continuous curve corresponds to a one with minimal $L_{\frac{1}{2}}$-norm.

In fig 7.5 is an example applying the local approach. It use the input location sequence as in the previous example (fig 7.3):

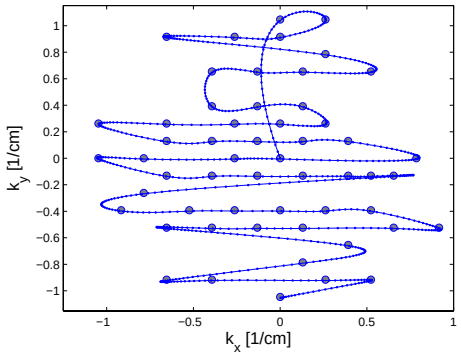


Figure 7.5: Result of $\boldsymbol{P}_{\text{kTrav}}$ using global approach

It ended up with 3.2ms to cover the 55 obligatory locations. (Nyquist grid with 1089 obligatory locations: 16ms)

Remark: In this example the global approach have better performance as the local approach. But we should be cautious to derive any general conclusions from it. Because of two reasons:

1. The performance depend strongly on the concrete distribution of the sampling locations to be covered. For some examples there are almost no difference in the final traveling duration.

2. The performance of the local approach is very strongly determined by the heuristic rule setting the velocities at the interior obligatory locations. The current rule is very simple and unoptimized. So the performance of the local approach has still strong margin to be improved.

7.5 The Validation Strategy

From the beginning of this thesis till now we have derived the original application task of spatial selective excitation to $\boldsymbol{P}_{\text{sparse}} + \boldsymbol{P}_{\text{kTrav}} + \boldsymbol{P}_{\text{RF-design}}$. We've also developed and implemented the according methods to approach these sub-problems. Along the way we have a number of critical points, which

are essential for a successful realization of our approach. Some of them are assumptions made by the model. Some of them are about the performance of our methods.

These critical points are the focuses of a systematical validation. We want to at first list these critical points for an overview. Then we design the validation strategy in such a way that we can check them separately as possible.

1. *The performance of solving the sparse approximation problem* $\boldsymbol{P}_{sparse}$

2. *The performance of our approach to* $\boldsymbol{P}_{kTrav}$
   —— whether the sparsity of the obligatory sampling locations can be sufficiently transferred to the actual acceleration of the k space traveling duration covering those locations.

3. *The STA assumption.*
   That the spin flip angle is sufficiently within the small tip angle regime is an essential assumption for the STA model. The whole linear representation formulism of the problem is based on this assumption.

4. *Sufficiently accurate* B_z *realization.*
   The $\boldsymbol{P}_{\text{SSEP}}$ assumes that the z-direction external magnetic field can indeed be sufficiently described by a spatially linear varying function, whose slope is the gradient waveform $g(t)$:

   $$B_z(\xi, t) = B_0 + \langle \xi, g(t) \rangle.$$

   But in the reality there are realization imperfections in the form of a) small additional non-linear spatial variation, $B_0 = \hat{B}_0 + B_0^{(\text{err})}(\xi)$, b) error by the gradient waveform realization: $g(t) = \hat{g}(t) + g^{(\text{err})}(t)$.

5. *The neglecting of relaxation.*
   It's a model assumption of $\boldsymbol{P}_{\text{SSEP}}$.

Now we separate our implementation and validation procedure in three major stages accordingly. We should be able to isolate the problematic part in case of any unexpected observation. These three stages are: *Sparse approximation stage*, *Bloch equation simulation stage* and *Phantom experiment stage*.

Stage-I: Sparse approximation :
Within the STA model, we validate the sparse approximation result. In this stage we can separately check the performance of our method solving $\boldsymbol{P}_{\text{sparse}}$, Checklist-No.1.

Stage-II: Simulation after $\boldsymbol{P}_{\mathbf{kTrav}}$, $\boldsymbol{P}_{\mathbf{RF\text{-}design}}$:

We take the selected sparse obligatory sampling locations and design a feasible k trajectory covering it. Taking the $4\mu s$ discretely sampled k trajectory sampling points, we solve the linear least squares problem $\boldsymbol{P}_{\text{RF-design}}$, which delivers the RF waveform. From the feasible k trajectory we calculate straightforwardly the feasible gradient waveform. We simulate the Bloch equation using the calculated gradient and RF waveform. We expect a simulation result very similar to the target profile and also the STA model prediction.

This stage separately validate the check-list No.2 and No.3.

Stage-III: Phantom experiment with MRI scanner :

After the validation via the Bloch equation simulation we replace the simulator with the real MRI scanner phantom experiment. This stage can validate the check-list No.4 and No.5.

A successful phantom experiment is a strong evidence for the practical applicability of our adaptive sparse concept in the praxis.

In section-7.6 we introduce the validation result of the sparse approximation level.

In section-7.7 we introduce the implementation and the validation of the Bloch equation simulation stage.

In section-7.8 we introduce the phantom experiment validation.

7.6 Validation Stage I: Sparse Approximation

In this section we will introduce the validation result of the sparse approximation stage. The target of this stage is to check the performance of the solving of $\boldsymbol{P}_{\text{sparse}}$, the checklist No.1.

The same heart shape target magnetization profile as previous is used, (fig 7.6 *top-left*). The complete ON Fourier dictionary is used, whose frequency domain sampling locations are determined after the Shannon Nyquist criteria 4.14. The $\boldsymbol{P}_{\text{sparse}}$ is approached by the AET algorithm. The error is measured by relative l_2-norm of the representation error.

The (fig-7.6-*top-right*) shows the error decaying behavior using the adaptive energy threshold algorithm. The x axis is the number of the participating atoms. The y axis is the representation error (RRMS). The error is represented in a logarithmic scalar.

As expected, the error decays very fast as the number of the participating k sampling locations increases. The point corresponding to the case of Nyquist grid fullsampling $(1089, 1.6\text{E}(-6))$, is far outside the plot range at the right side. In the table 7.2 are some representative data:

	TOL	Number of Atoms	Representation Error
A).	*full Nyquist*	1089	1.6E(−6)
B).	1%	431	0.9987%
C).	5%	121	4.96%
D).	10%	65	9.85%
E).	15%	43	14.87%

Table 7.2: The sparse approximation result.

The represented profiles and their participating k sampling locations regarding different accuracy requirement are shown in the bottom columns of (fig-7.6).

By *Nyquist fullsampling* (fig.-7.6-block-A) one gets exactly the target profile, as expected. With *TOL = 1%* one gets essentially the same as the Nyquist fullsampling result. But one has a significant reduction of the k sampling points, from 1089 to *431*.

The number of atoms reduces further by TOL= 5% to *121*. The discrepancy is restricted. The 5% is a very typical user accuracy requirement. Since the Hardware realization accuracy is also typically about 5%, it usually won't bring in more quality benefit at the end for the accuracy better than 5%.

Further to 10% and 15% the number of the necessary atoms get further reduced to 65 and 43 respectively. At the same time one also starts to observe significant quality reduction. Especially in the TOL=15% case. The discrepancy appears in the forms of the blurring effect in the heart figure. The Gibbs ring at the outside can also be observed, which means we really cut into the area where the frequency coefficients are *not* small.

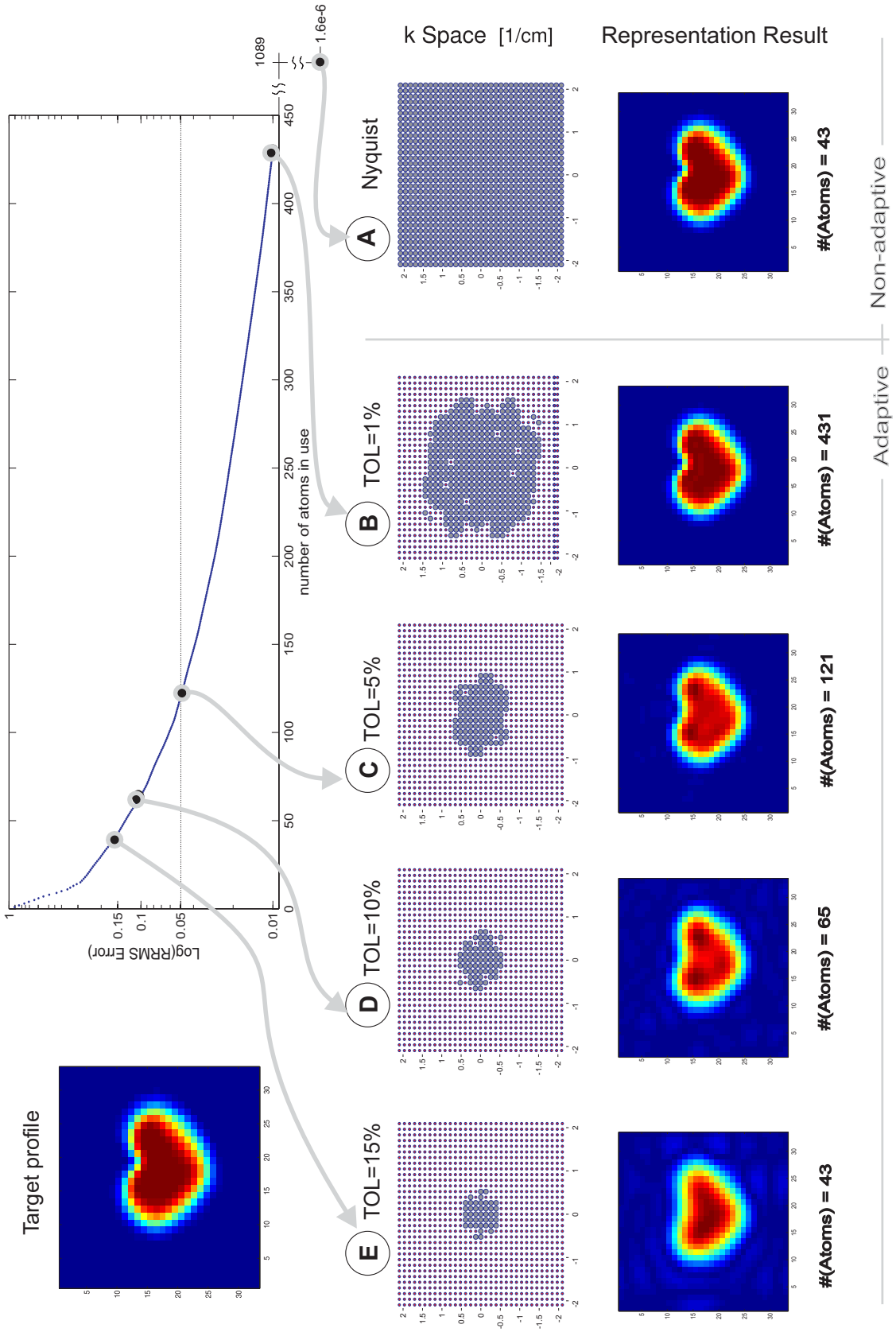


Figure 7.6: The validation of the sparse approximation algorithm

7.7 Validation Stage II: Bloch Equation Simulation

In this stage we do the following four steps to obtain the two control waveforms, G and RF, and perform in turn the simulation validation:

1. *Solving* $\boldsymbol{P}_{kTrav}$.
   The $\boldsymbol{P}_{\text{kTrav}}$ is approached as described in (§-7.4). (The local approach is used here.)

2. *Solving* $\boldsymbol{P}_{RF\text{-}design}$.
   The actual RF waveform is obtained by solving the according linear least squares problem regarding the physically feasible k space trajectory with the discrete sampling distance of $4\mu s$. The LSQR algorithm [PS82] is used here.

3. *Calculating the gradient waveform* $k(t) \mapsto g(t)$.
   The actual gradient waveform is obtained by the derivative of the feasible k space trajectory $k(t)$.

4. *Validation via the Bloch equation simulation.*
   The obtained $g(t)$ and RF(t) will be used as the input waveform of the Bloch equation simulation for validation.

7.7.1 $\boldsymbol{P}_{\text{kTrav}}$ and $\boldsymbol{P}_{\text{RF-design}}$

(Fig-7.7) demonstrates the paradigm concretely for the single channel example case-D: (TOL=10%,$R = 7.1$).

After selecting out the sparse obligatory sampling locations $\boldsymbol{\kappa} := \kappa_1, \cdots, \kappa_L$ in the frequency domain, we calculate a feasible k space traveling trajectory $k(t)$ to cover $\boldsymbol{\kappa}$ and subject to the feasibility constraints 4.15. it is the problem ($\boldsymbol{P}_{\text{kTrav}}$). We approach it as discussed in section 7.4.

To see how the actual traveling duration correlates with the sparsity of the obligatory sampling locations from the sparse approximation stage, we plot the k-traveling-duration vs. the number of the obligatory sampling locations to be covered.

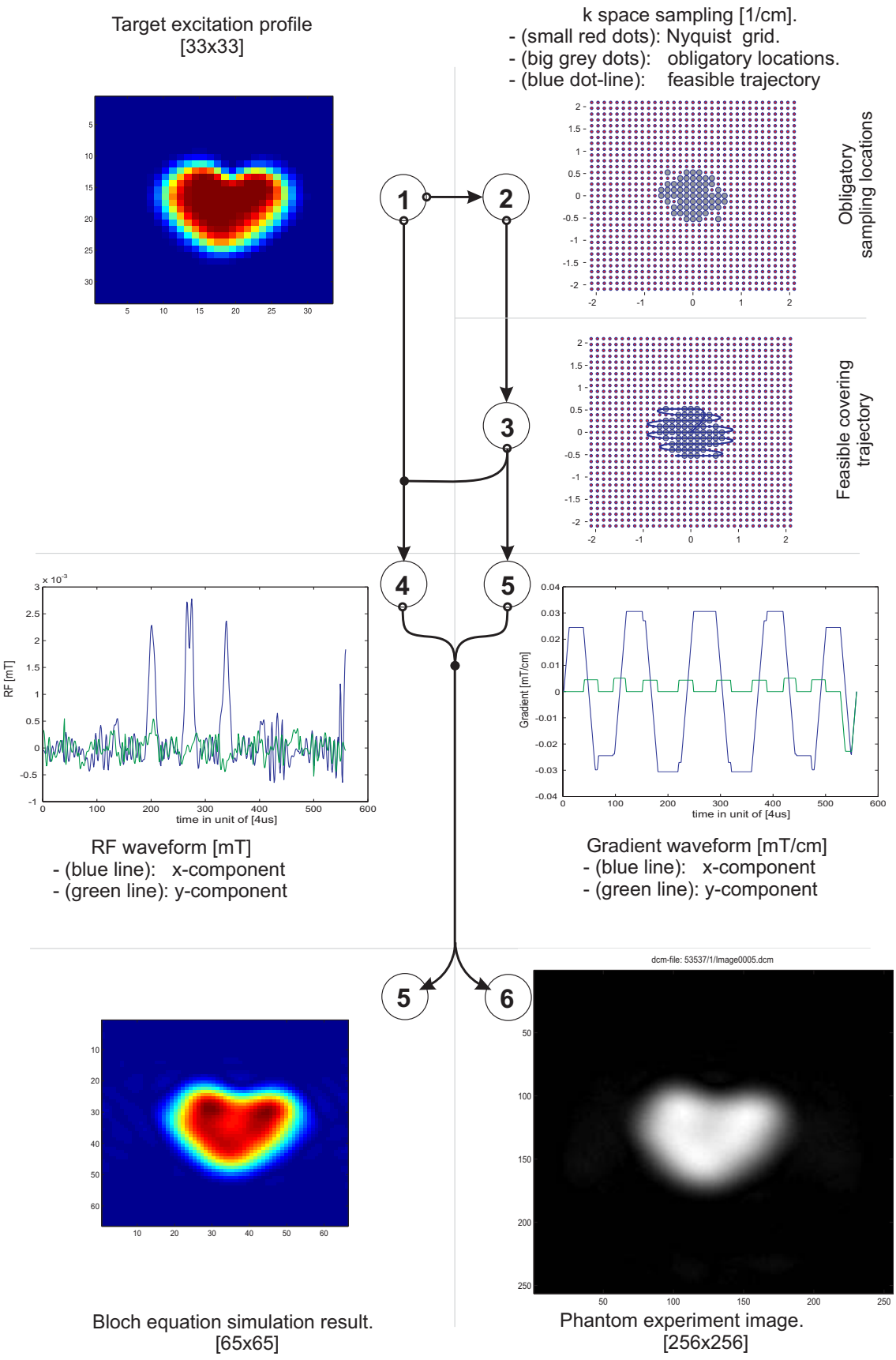


Figure 7.7: The example case-D, single channel setting

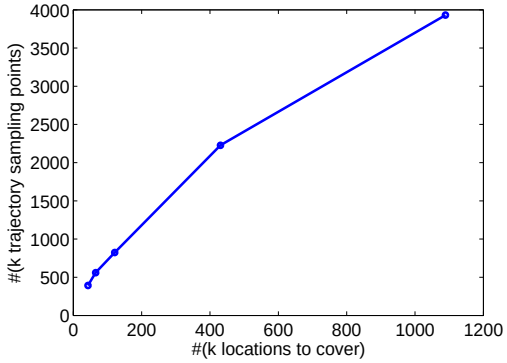


Figure 7.8: #(Obligatory locations) vs. actual traveling duration
x: #(obligatory sampling locations) vs. y: #(feasible k trajectory sampling-points with $4\mu s$ sampling distance)

We see it is a very linear trend. This means the assumption that the undersampling effect can be sufficient transferred to the final acceleration effect is valid.

The resulting feasible trajectory $k(t)$ is in turn discretely sampled with a fixed sampling rate of $[4\mu s^{-1}]$ due to the concrete hardware realization circumstance. At the end the Fourier vectors with those discrete sampling points as frequency parameters are the indeed participating vectors to represent the target profile.

Therefore, the best achievable representation is calculated by solving the linear least squares problem using all the participating Fourier vectors, $\boldsymbol{P}_{\text{RF-design}}$.

Mostly it ends up with more k sampling points than the number of the target profile pixels. One hat *formally* an under-determined linear least squares problem. By just applying the Moore-Penrose pseudo inverse one get a solution with the least $l2$-norm among all the other possible solutions fulfilling the least squares requirement. The l_2-minimal solution has the physically meaning of a minimal global energy deposition (SAR: Specific Absorption Ratio). Minimizing the energy deposition is exactly what we want beside acceleration. Therefore the Moore-Penrose pseudo-inverse is just the natural choice for this formal underdetermined linear least squares problem.

Remark: The under-determinedness is only *formal*. Because additional sampling points are on the connecting path between the neighboring Nyquist grid members. The according Fourier vectors are a linear dependent vector set.

7.7.2 Bloch equation simulation results

We generate the gradient waveform $g(t)$ directly as a differential form of $k(t)$.

We take $g(t)$ and RF(t) as input of the Simulation for the relaxation free Bloch equation (4.1). If there would be any inconsistency due to the STA assumption, it would be shown by the simulation result of the Bloch equation. Because the simulation of the Bloch equation is independent from the STA model.

Since at the end we use more sampling points (atoms) than the original obligatory sampling location, the actual excitation quality will usually get improved from the representation quality from the sparse approximation stage. The fig-7.9 is the plot of the user TOL vs. the actual error predicted by Bloch simulation.

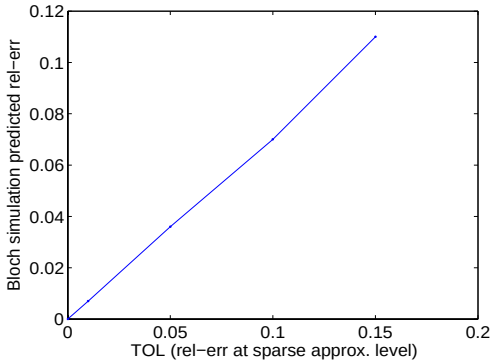


Figure 7.9: The user tolerance for the sparse approx. algorithm vs. the actual error predicted by Bloch simulation

We see a very linear behavior. This means the end quality is good pridictable by multiplying an empirical coefficient to the user tolerance.

To check any unwished oscillation within the pixels size regarding the target profile resolution, the Bloch simulation are performed with a finer spatial resolution as the target profile. Because the later imaging process is at a much finer resolution as the current target profile resolution. If we don't perform such a finer resolution check. It could happen that the actual image have strong oscillations at higher resolution, while in the lower target profile resolution level very thing looks fine. (Usually the excitation target profile resolution are set much lower than the final imaging resolution to reduce the numerical burden of the waveforms design. The reason one can do that is that most excitation target profiles have information low dimensionality nature in frequency domain. 33×33 or 65×65 are empirically confirmed good resolutions.)

We measure the quality of the result by taking the relative-error of the l_2-norm of the difference of the target profile and the Bloch equation simulated profile.

The plot of (fig-7.10-*top-right*) shows the actual error vs. excitation duration behavior.

In table-7.3 is some representative data. The conventional paradigm of Nyquist full-sampling ends up with a feasible k trajectory of 15.7ms (3931 sampling points with a sampling distance of $4\mu s$). We take this value as reference for calculating the reduction factor $\boldsymbol{R}$ to measure the acceleration:

$$\boldsymbol{R} := \frac{\text{Excitation duration [ms]}}{15.7\,\text{ms}}.$$

	TOL	Excitation Duration	Reduction Factor	Excitation Error
A).	*full Nyquist*	15.7ms	$R \equiv 1$	1.0E(−6)
B).	1%	8.9ms	$R = 1.8$	0.7%
C).	5%	3.3ms	$R = 4.7$	3.6%
D).	10%	2.24ms	$R = 7.1$	7%
E).	15%	2.0ms	$R = 10.5$	11%

Table 7.3: Some Bloch simulation data.

The excitation profiles and their according k space trajectories regarding different accuracy requirement are shown in the bottom blocks of (fig-7.10).

Through the validation via Bloch equation simulation we see that the sparsity of the k space obligatory sampling locations can be sufficiently transferred to the desired benefit of actual acceleration.

By high reduction factor, it has the same trend of the blurring and Gibbs ring effect, which has already been observed in the sparse approximation stage.

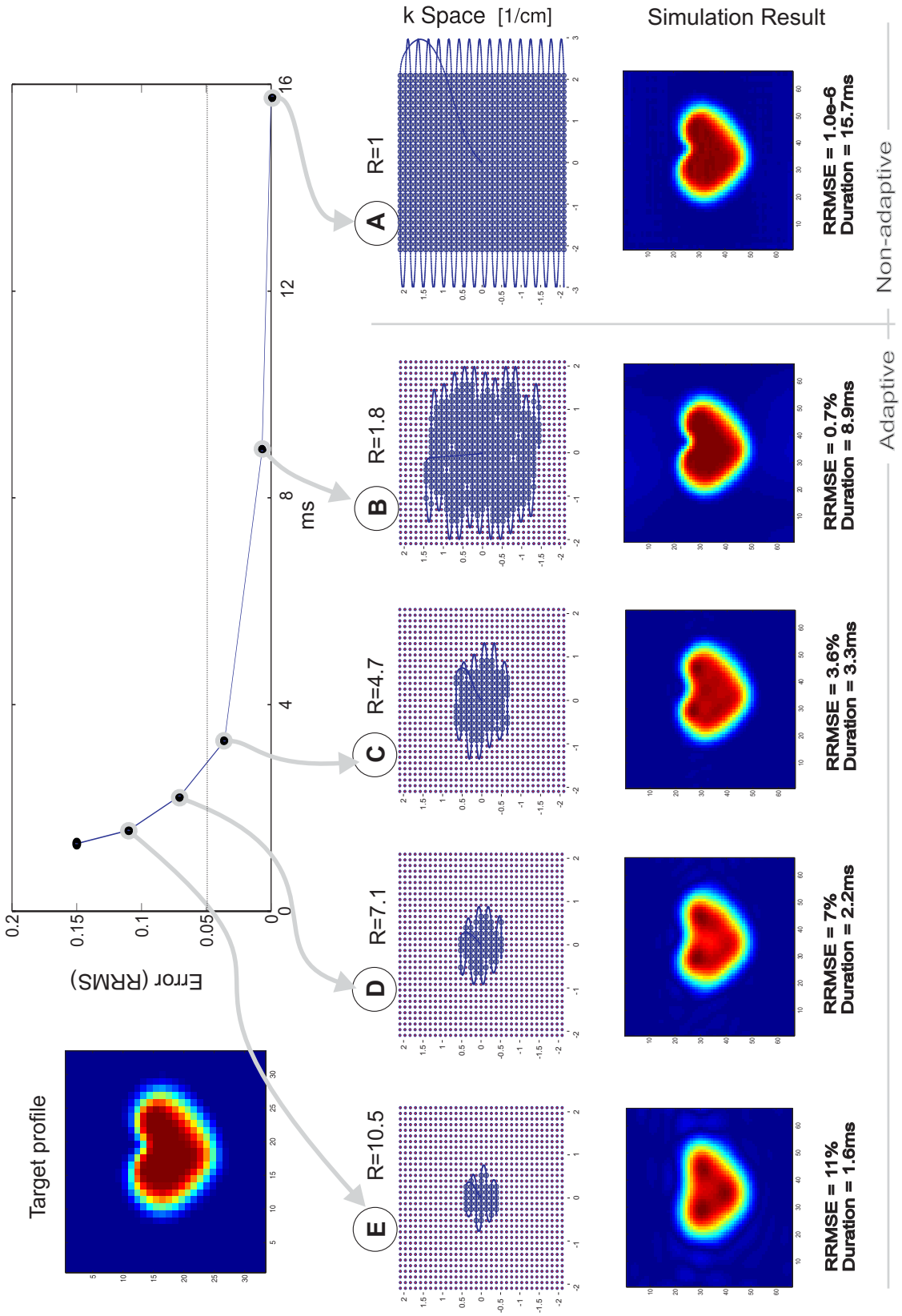


Figure 7.10: The validation via the relaxation-free Bloch equation simulation

7.8 Validation Stage III: Phantom Experiments

In the previous two stages, I. sparse approximation and II. Bloch equation simulation, we validated the critical points No. 1, 2, 3 in the checklist (§7.5). In the section we introduce the validation stage of phantom experiment. The target of this stage is a further proof of principle of the practical applicability of the proposed method. Concretely, it aims to validate the checklist No. 4, *whether the hardware realization of* $B_z(\xi, t)$ *is sufficiently accurate*, and the checklist No. 5, *whether the relaxation process is negligible*. If we can obtain similar excitation images as the Bloch equation simulation predicted, then we can say both the checklist points 1 and 2 are fine. It is strong evidence that the proposed method is practically applicable.

We did our experiment with a GE MRI 3T scanner. We use a water phantom. It has a ball shape with diameter about 15cm. In fig-7.11-*right* is the phantom.

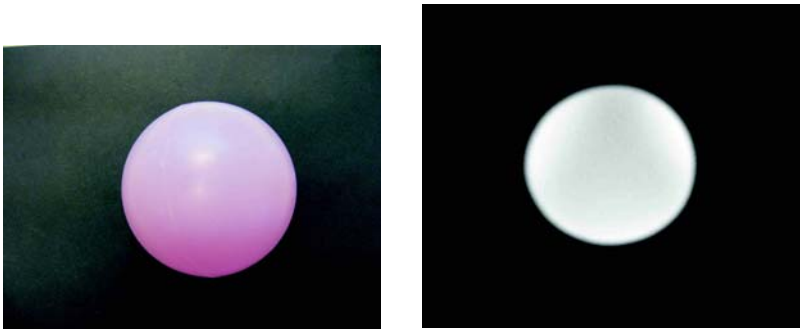


Figure 7.11: The phantom (*left*) and its no-selective excited MRI image (*right*)

Without any selectivity by the excitation, everywhere in the phantom the hydrogen atom in the water molecule will be excited more or less homogeneously. One will see a disk at a xy-plane 2D image (fig-7.11-*right*).

We show the images from the phantom experiment together with the according images predicted by the Bloch equation simulation from the last section. As the convention in the MRI literature, the MRI scanner experiment image are shown in gray-scale and the simulated images are shown in color.

An interesting observation is of the comparison between phantom image A and B. The case B with user tolerance 1% and is about half of the excitation duration as the case A of the conventional paradigm using the Nyquist

full-sampling. The phantom image of the Nyquist full-sampling case shows however some artifact which is not observed in the validation stage of Bloch equation simulation. It has a less good quality than the phantom image of case B. One probable explanation for that is the too long excitation duration of the Nyquist full-sampling solution. Under this long duration the hardware imperfection accumulates to excitation artifact. The relaxation of the spins can also be the reason. Neither of these two facts are included in the Bloch equation simulation. That's why it wasn't observable by the validation via simulation.

This observation exactly demonstrated the needs for accelerating the excitation process and the benefit of the proposed adaptive paradigm.

Except that, the phantom experiment images match the simulation prediction very good. They also show the same excitation quality change of blurring and Gibbs-ring as predicted by the Bloch simulation. (Due to the printing quality, the Gibbs ring effect may be better observable in the electronic version (attached CD-Rom) of this thesis than in the printed version.)

The matching of the simulated images and the phantom experiment images is a strong evidence that the proposed method is not only theoretical consistent but is also applicable in the praxis.

Remark: *A flaw by the experiment.* The phantom images are a little bit smaller than the simulation prediction. The reason for that flaw is a wrong scaling coefficient of the gradient waveform by the phantom experiment. Since $k = \int g$, an additional scaling coefficient $g \leftarrow c \cdot g$ cause the scaling of the k space sampling grid. A general scaling of the k space locations, but not their coefficients cause a scaling of the images in the spatial domain. But the content of the images will not be change no matter with or without this scaling coefficient c. For the aim of proof of principle, this flaw of the scaling factor doesn't has any influence on the conclusion. To make the comparison easier, the phantom images are scaled such that both profile shape are of similar size.

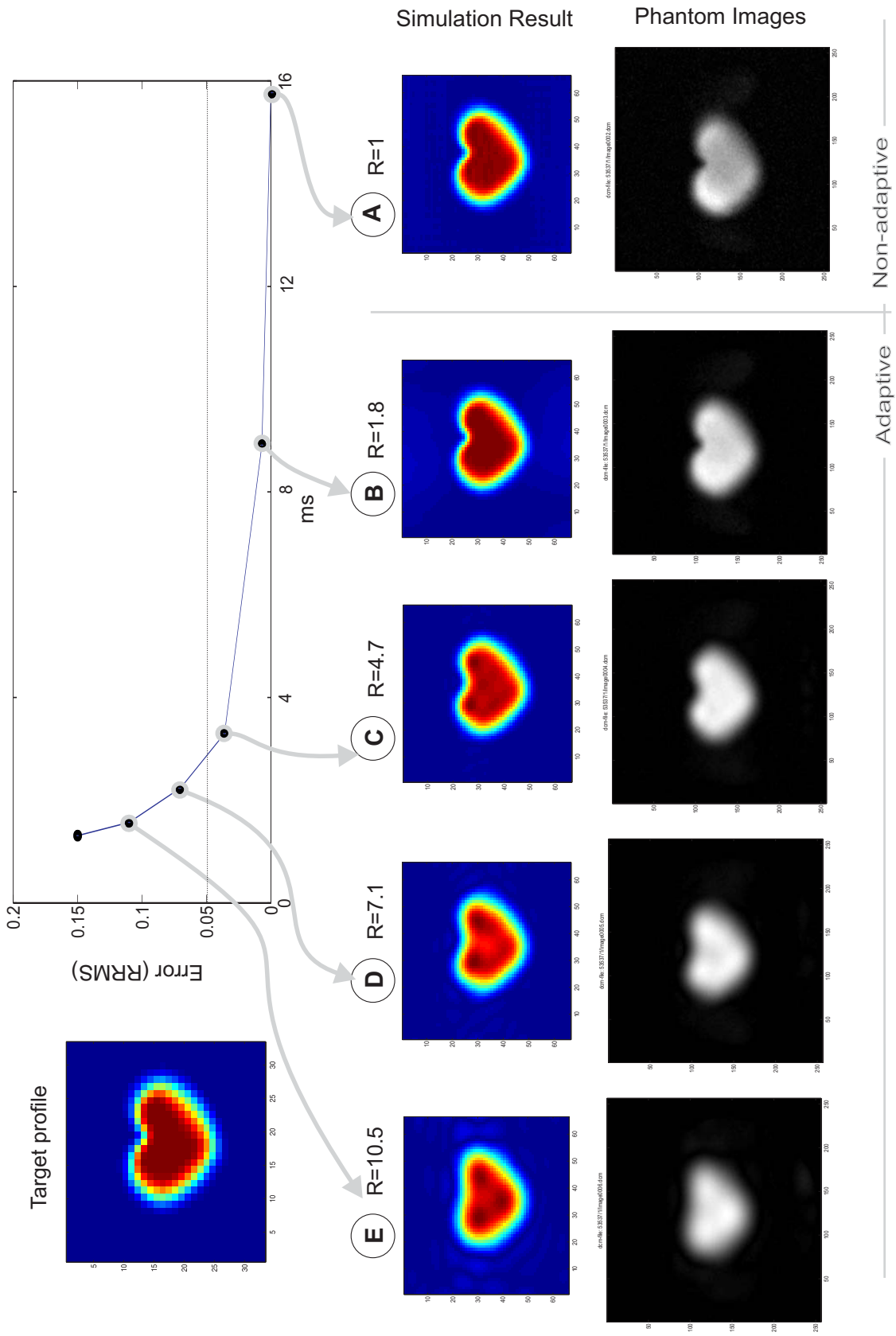


Figure 7.12: The validation results via the phantom experiments

7.9 The Adaptivity

After implemented and successfully validated the proposed adaptive sparse method, we devote this section to demonstrate the adaptive capability of this method.

Concretely, the adaptive capability includes to components:

1. The adaptivity to the user given accuracy requirement "TOL".
   It means the capability of automatically adjust the solution depending on the accuracy requirement. (fig 7.6 *top-right*), (fig-7.10-*top-right*).

2. The adaptivity to the user given target excitation profile.
   It means, depending on the concrete target profiles, especially their complexity levels, the proposed method automatically determines how many and where to sparsely sample the k space.

The following four example cases together clearly demonstrate the adaptive character of the proposed method.

The three target excitation profiles (fig-7.13) represent three different information complexity of the target profiles.

A). Low complexity (*fig-7.13-A.*) The sensitivity profile of a localized RF transmit coil.
This is the sensitivity profile of a RF transmit coil for a ball-shape water phantom. (The RF transmit sensitivity has been discussed in the §-4.2.) The area outside the phantom has no signal, because the hydrogen atom distribution in this area is zero. We call it the don't-care region. Only the area within the circle is interesting for us. The sharp boundary of the circle shape is not the core part of the information. It can be removed by a smooth extrapolation from the interior data. This profile represents the target profile with very low information complexity. This level of complexity corresponds to the application type of *"B_1 homogenization"*.

B). Intermediate complexity: (*fig-7.13-B.*) The heart figure.
The heart profile represent the intermediate complexity level of the target profile. it is a representative profile for the application type of small region excitation.

C). High complexity: (*fig-7.13-C.*) The Mona-Lisa profile.
This profile of Mona-Lisa portrait represents an extreme case of the profiles with very high information amount. The investigation of this kind of fancy profiles is rather of the theoretical interest. This kind of

extreme complex profiles are still unlikely to really be applied in the praxis.

All the profiles are represented in a spatial resolution of 33×33, which is a very widely used resolution for the spatial selective excitation task.

In the conventional non-adaptive paradigm, the k space trajectories are based on the full-sampling of the Nyquist grid. Since all the three profiles have the same spatial grid, they have all the same Nyquist sampling grid. Therefore one ends up with the same k trajectory and thus the same excitation duration. The obvious different complexity level of the profile content and the different user accuracy requirement are not taken into account.

In contrast, the proposed adaptive method is constructed to automatically recognize the complexity and the accuracy requirement and accordingly design the excitation as short as possible.

The comparison of A, B and C demonstrates that the adaptive method delivers automatically solutions with differently long excitation durations tailored to the different information complexity of the target profile. The comparison of C1) and C2) shows again the adaptivity to the accuracy requirement as already be shown in (fig 7.6) and (fig-7.10).

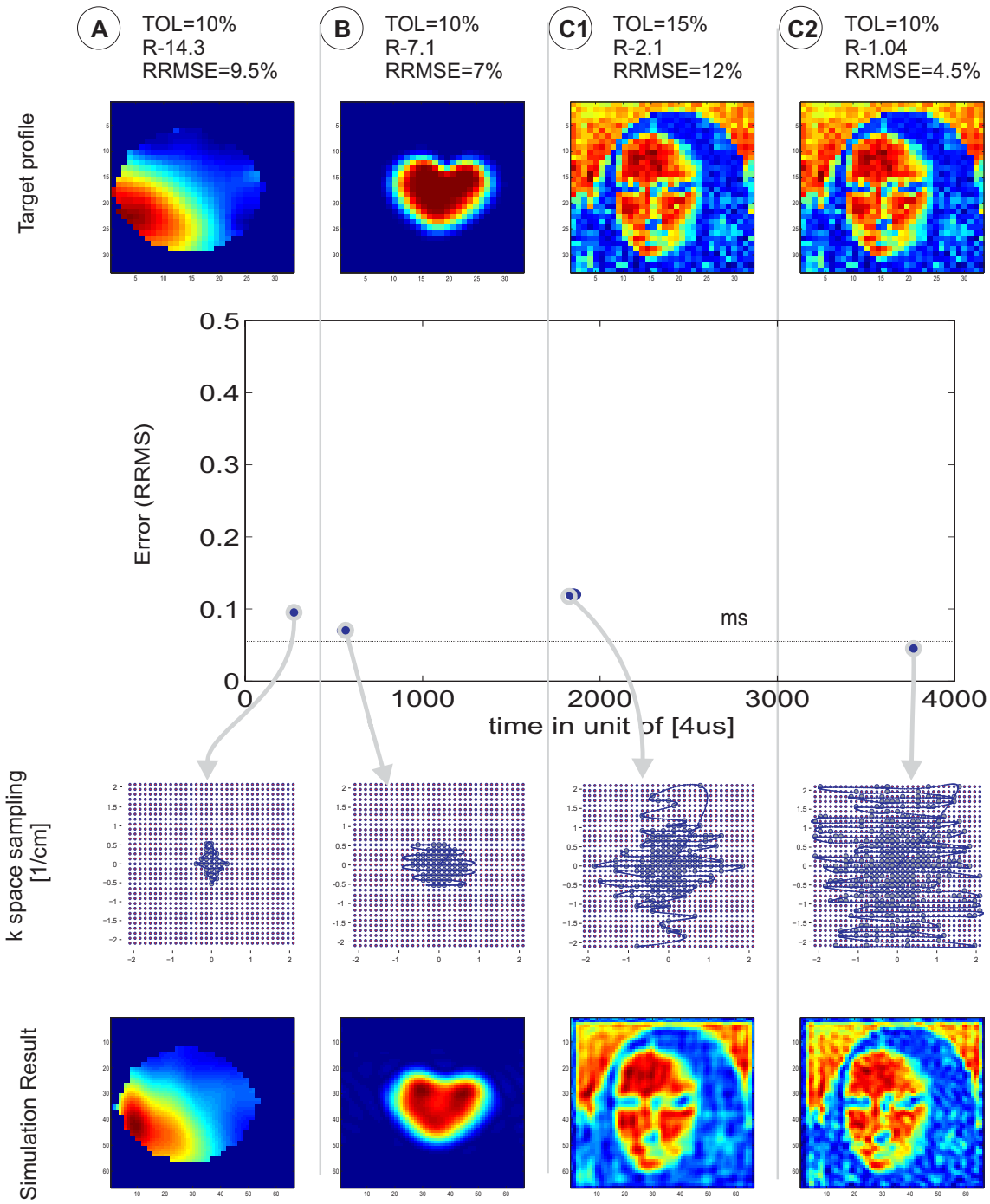


Figure 7.13: The adaptivity demonstration

Chapter 8

Application Setting 2: Sparse Parallel RF Transmit

8.1 Recall the Problem: Adaptive Parallel RF Transmit

In chapter 4 we discussed the parallel transmit techniques. We recall the application setting.

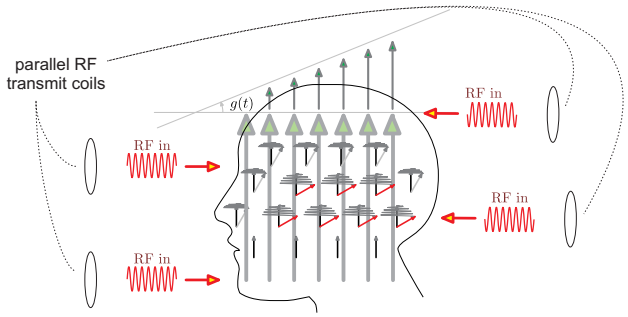


Figure 8.1: Schematic for the parallel transmit

With the STA model, the task of $\boldsymbol{P}_{\text{SSEP}}$ using parallel transmit setting can be reduced to two subproblems: $\boldsymbol{P}_{\text{kTrav}}(\boldsymbol{\kappa}_{\text{Nyq}})$ (4.15) and $\boldsymbol{P}_{\text{RF-design-(pTx)}}$ (4.22).

Also under the STA model, the proposed adaptive paradigm divides the parallel-$\boldsymbol{P}_{\text{SSEP}}$ to three subproblems $\boldsymbol{P}_{\text{sparse-pTx}}$ (4.27), $\boldsymbol{P}_{\text{kTrav}}$ (4.15) and $\boldsymbol{P}_{\text{RF-design-(pTx)}}$ (4.22). The $\boldsymbol{P}_{\text{sparse-(pTx)}}$ can be seen as a special form of sparse approximation problem. The $\boldsymbol{P}_{\text{kTrav}}$ is formally the same as the one

for the single channel setting (Chapter 7). However in the parallel RF transmit setting the input obligatory sampling locations are more irregular. The $\boldsymbol{P}_{\text{RF-design-(pTx)}}$ remains essentially the same as in the non-adaptive parallel transmit setting. In can be solved using standard numerical software, e.g. LSGR. [PS82] [ZWS+07]

Like for the single channel setting, we follow the pragmatic strategy for the first two subproblems: first concentrate on solving ($\boldsymbol{P}_{\text{sparse-pTx}}$) as good as possible, then solve ($\boldsymbol{P}_{\text{kTrav}}$) sufficiently good such that the achieved undersampling enhancement can be transferred to the final desired effect of acceleration.

The structure of this chapter is similar to the last chapter.

In the section 8.2 we discuss applying subspace-OMP approaching ($\boldsymbol{P}_{\text{sparse-pTx}}$).

Since the approach of $\boldsymbol{P}_{\text{kTrav}}$ are the same as in the single channel setting, we just point to the section 7.4 in the last chapter.

In section 8.3 we introduce the very similar validation strategy as in the single channel setting.

In section 8.4 we validate the undersampling-enhancing property of the subspace OMP algorithm in the STA model.

In section 8.5 we validate the final acceleration effect of our method via simulation of the relaxation free Bloch equation.

In section 8.6 we validate the praxis applicability of our method via the phantom experiment in a GE 3T MRI scanner.

8.2 The Synchronized Sparsity of $\boldsymbol{P}_{\text{sparse-pTx}}$ and the Subspace OMP Method

(The content of this section is partially redundant with the discussions in §-4.3 and Chapter-6 — a kind of summary reorganized from an application perspective. This is mainly for the MRI reader who skipped the Part-II.)

8.2.1 Dictionary: Gabor type dictionary

At first for the choosing the appropriate candidate set of k sampling locations, we follow the similar arguments discussed in the last chapter for the single channel setting (§-7.2). We choose the Nyquist sampling grid as the candidate set of the k sampling locations.

The resulting dictionary consists of the atom vectors $d_k^{(q)} \in \mathbb{C}^N$ with two parameters k and q.

$$d_k^{(q)} = \text{Diag}(s^{(q)}) \cdot e^{ik\xi}$$

Where $\xi \in \mathbb{C}^N$ is the vector for the spatial coordinates of the pixels, $s^{(q)} \in \mathbb{C}^N$ the q-th spatial sensitivity profile of the RF waveform.

So the atom vector $d_k^{(q)}$ turns to be a spatial Fourier harmonics with the frequency parameter k modified by a spatial weighting profile $s^{(q)}$. This remember us to the so-called *Gabor dictionary*.

Our dictionary can be seen as a special form of the Gabor dictionary with two parameters: the frequency parameter k, and the spatial weighting function parameter q.

8.2.2 What's the problem if apply normal sparse approximation method

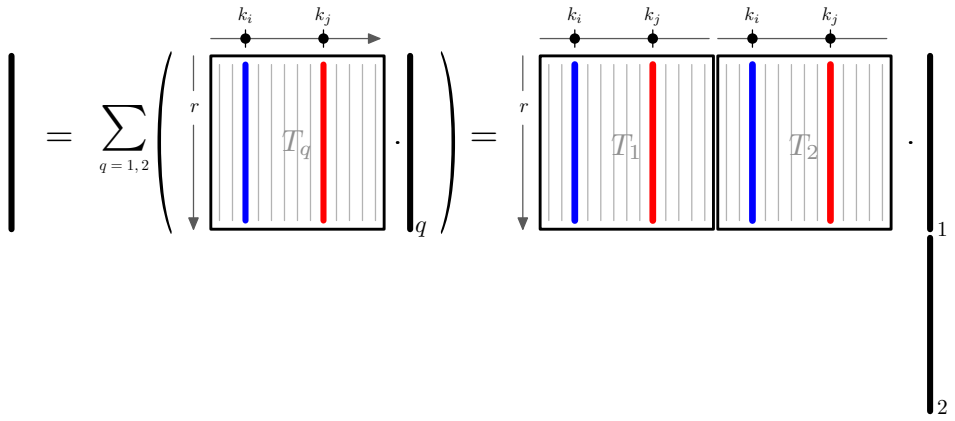
The $\boldsymbol{P}_{\text{sparse-pTx}}$ looks very like the sparse approximation problem using Gabor type of dictionary, which can be approached using standard heuristic method like OMP or L_1-minimization. However the $P_{\text{pTxSparse}}$ has a essential difference on the sparsity requirement compared with the normal sparse approximation problem.

The conventional sparse approximation problem are discussed in Chapter-5. Its the objective is to use as less atom vectors as possible from the dictionary to represent the target vector $\boldsymbol{p}$.

However the objective of $\boldsymbol{P}_{\text{sparse-pTx}}$ is to use as less frequency parameters k as possible. The spatial weighting function parameter q is not included as a sparsity objective due to the application physical setting.

We call the later sparsity requirement the synchronized sparsity, to emphasis from the application side that all the RF transmit channels share the same k space sampling pattern. Sometimes we also call it the subspace sparsity, to emphasis from the sparse approximation methodology side that here one doesn't pursuit the sparsity of the individual atom vectors, but pursuit the subspaces spanned by a group of atoms with the same frequency parameter k.

It's obvious that the minimal participating atoms vectors (k, q) is not the same as the minimal participating frequency parameters (k). This can be demonstrated more clearly through the following schematics:



Since both sparsity objective are not the same. The in chapter-5 introduced standard numerical methods are in general not directly applicable for solving $\boldsymbol{P}_{\text{sparse-(pTx)}}$.

However if we extend the concept of vector atoms into a general concept of subspace atoms, one can modify the greedy method to meet the synchronized sparsity requirement.

8.2.3 The idea: subspace atoms

The special sparsity objective of $\boldsymbol{P}_{\text{sparse-(pTx)}}$ can also be reformulated in the following way. If one picks up a k sampling location, one has simultaneously C available to linearly represent the given target vector. Here all C vectors have the same frequency parameter k. To compare the importance of the different k sampling locations is nothing else than to compare the importance of C according vectors.

This view brings us straightforwardly to the idea of subspace atoms. In the normal sparse approximation problem, the element atoms are vectors, which compete each other for an economic linear presentation. By replacing the vectors with the C-dimensional subspaces, the special sparsity requirement will be guaranteed.

We call this special form of sparse approximation the subspace sparse approximation problem.

In our concrete context of Gabor dictionary, the subspace sparse approximation varies only the frequency parameter, instead of varying both frequency and spatial weighting parameters. If a frequency parameter is picked up, all the spatial weighting parameter combinations will be taken.

8.2.4 The subspace OMP method

The extension of the normal OMP algorithm to the subspace concept is quite straightforward. The essential step is how to "compare" the atoms. In vector atom case, it's compared by their vector inner product with the residual vector. This vector inner product is nothing else than the projection of the residual vector to the one dimensional subspaces spanned by the according vector atoms. Its equivalence in the subspace case is to project the residual vector to the multidimensional subspace, and weighting the projection vector by their Euclidean norm. Instead of the vector inner-product, one has the more general form of a small linear least square problem of size $N \times C$, to project the residual to the subspace.

Algorithm 5 Subspace OMP

1: {Initial.}

$$l \leftarrow 1, \quad \Lambda \leftarrow [\,], \quad \mathrm{res}_0 \leftarrow \boldsymbol{p}$$

2: {Find the next best index}

$$\lambda_l = \underset{k}{\operatorname{argmax}} \left(\| \hat{\mathcal{P}}_{\boldsymbol{s}_k}(\mathrm{res}_{l-1}) \|_2 \right)$$

3: {Update the index vector}

$$\Lambda_l = (\Lambda_{l-1}, \lambda_l)$$

4: {Update the coefficient vector, the residual}

$$\boldsymbol{b}_l = \underset{\boldsymbol{b} \in \mathbb{C}^l}{\operatorname{argmin}} \left(\| \boldsymbol{p} - \phi_\Lambda \cdot \boldsymbol{b} \|_2 \right)$$

$$\mathrm{res}_l := \| \boldsymbol{p} - \phi_\Lambda \cdot \boldsymbol{b}_l \|_2$$

{Check for break}

5: **if** $\| \mathrm{res}_l \|_2 > \mathrm{TOL}$ **then**
6: $l \leftarrow l + 1$; $\rightarrow$ Step 2
7: **else**
8: **return** Λ, $\boldsymbol{b}$ with success!
9: **end if**

8.2.5 The theoretical aspect

In chapter-6 we discussed the theoretical two aspects for the subspace sparse approximation, namely the uniqueness condition of the subspace sparse ap-

proximation and the optimality condition of the subspace OMP algorithm.

Unfortunately, the presented results can still not offer much help to in the praxis. There are several reasons for the current restriction for the practical usage.

First of all, it's a general nature of the current theoretical results in sparse approximation. The investigation follows mostly the worst case scenario and ends up with a most likely too pessimistic boundary estimation.

This is especially the case, if the used dictionary doesn't have sufficient incoherence. Our application case has exactly this problem. The physical model restricts the choose of dictionary in basic on Fourier harmonics. The spatial weighting function, the spatial sensitivity profile of the RF waveform transmit coils, has normally not enough orthogonality due to the physical restrictions by the hardware design.

This leads to the issue in the praxis that one can calculate the rigorous condition for both uniqueness and optimality, but the sparsity requirement of the guarantee mostly ends up at some unrealistic small number. The theorem can only make concrete statement for some very special cases with extreme high sparsity.

The second reason is specially due to the subspace extension. In the uniqueness theorem for subspace sparse approximation, one must exclude a intersection region $\hat{H}(\mathfrak{D})$ to make a rigorous statement. However this $\hat{H}(\mathfrak{D})$ is very unpractical to calculate.

Thus the practical meaning of the statement is essentially reduced to an unrigorous statement: For an arbitrary vector from $\mathbb{C}^N$, it's unlikely to be lying in the $\hat{H}(\mathfrak{D})$ because the $\hat{H}(\mathfrak{D})$ is a null-set in the $\mathbb{C}^N$.

Third reason is rather a temporal one due to the current state of the investigation. The current two theorems we introduced are rather for the subspace sparse *representation* than the subspace sparse *approximation*, because we discussed only the special case for zero user tolerance, $\varepsilon = 0$. In contrast, in praxis we have quite large tolerance, about 5%. The restriction to the zero tolerance makes the uniqueness and optimality condition unnecessary strict.

In the praxis, the heuristic experiments shows promising result. It shows that the subspace OMP method can strongly enhance the sparsity. In the next three sections we will introduce the implementation and the validation of the proposed method.

We follow the same implementation and validation strategy as for the single channel setting. (§-7.5)

8.3 Validation Strategy for pTx Setting

The validation strategy for the adaptive sparse parallel RF transmit is very analog to the validation strategy of the single channel setting (§-7.5).

We at first give a checklist for the critical points for a successful realization of the adaptive sparse parallel transmit:

1. *It checks the performance of solving the subspace sparse approximation problem $\boldsymbol{P}_{sparse\text{-}pTx}$ using subspace OMP method.*

2. *It checks the performance of our approach to $\boldsymbol{P}_{kTrav}$*
   It checks whether the sparsity of the obligatory sampling locations can be sufficiently transferred to the actual acceleration. The adaptive pTx delivers usually sparser and at the same time more irregular distributed obligatory sampling locations. Therefore the sparsity-acceleration transfer is more critical for the parallel transmit setting than in the single channel case.

3. *The STA assumption.*
   As for the single channel setting, one needs to check the availability of the model assumption of small tip angle regime.

4. *Sufficiently accurate B_z realization.*
   The spatially linear varying B_z is a model assumption in the problem $\boldsymbol{P}_{\text{SSEP}}$. The problematic is basically the same as for the single channel setting (§-7.5). However, since the adaptively calculated k trajectories by the parallel transmit setting are usually more irregular (non-symmetric in the k space) than the trajectories by the single channel setting, one expects the task of a sufficiently accurate hardware realization of the according strongly irregular gradient waveform to be more demanding.

5. *The sensitivity profiles*
   One factor one doesn't have by the single channel setting is the sensitivity profiles of the RF transmit channels. A correct sensitivity profiles acquisition is essential for using parallel transmit setting. The concrete methods to acquire the transmit sensitivity profiles is a seperate topic, which is beyond the scope of this thesis. Here we just take them as given. However since the successful realization of the parallel transmit concept presume a sufficiently accurate knowledge of the sensitivity profiles, we should keep in mind that an incorrect sensitivity profile could cause mismatch between the simulation and the phantom experiment result.

6. *The neglecting of relaxation.*
   As in the single channel case, it's a model assumption of $\boldsymbol{P}_{\text{SSEP}}$.

Similar to the single channel case, we divide the validation in three stages: *Subspace sparse approximation stage*, *Bloch equation simulation stage* and *Phantom experiment stage*.

In section-8.4 we discuss the subspace sparse approximation stage. The validation target is the checklist No.1.

In section-8.5 we the Bloch equation simulation stage. The validation target of this stage is the checklist No.2 and 3.

In section-8.6 we the phantom experiment validation stage (Checklist No.4, 5 and 6).

8.4 Validation Stage I: Subspace Sparse Approximation

The goal of this stage is to check the performance of the subspace OMP method to approach $\boldsymbol{P}_{\text{sparse-pTx}}$.

We use a realistic setting of six independent RF transmit channels and their sensitivity profiles of a ball phantom. Fig-8.2-*left* shows the originally measured sensitivity profiles of the six transmit channels.

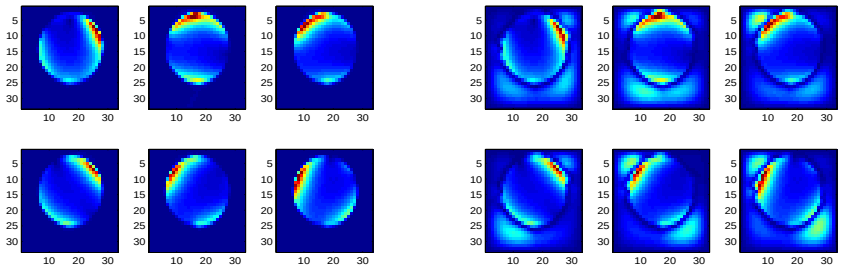


Figure 8.2: The sensitivity profile
(*left:* the measured sensitivity (absolute value is plotted here). *right:* sensitivity profile extrapolated into the don't-care region (decaying to zero))

Since one get only signals from inside the ball. We have no information about the RF sensitivity outside the ball region. But we also don't care about this region. Because the excitation process only happens inside the object.

The border between the phantom ball and the don't-care region is a sharp discontinuity jumping from the interior sensitivity value to zero outside. This discontinuity is however not a part of the physical property of the sensitivity, but purely caused by the lack of information from the don't-care outside region.

On the other side, we know the smoothness in the spatial domain determines how compact is the distribution in the frequency domain. Any discontinuity in the spatial domain causes additional sampling points in the frequency domain.

So we should eliminate the unnecessary discontinuity caused by the phantom boundary. We do it by smoothly extrapolate the original interior sensitivity data to zero at the boundary of the FOV in spatial domain. In the fig-8.2-*right* the extrapolated sensitivity profiles are shown We will use the extrapolated sensitivity profiles instead of original discontinuous ones from now on.

How to extrapolate the original data to eliminate the unnecessary discontinuity has large influence on the resulting sparisty of the solution. It is not a trivial task, especially because the original date varies very strongly in the region near the phantom border. Hence it is not easy to "tell" the algorithm from where on is the don't-care region to start the extrapolation.

We believe the current extrapolation is still not the optimal one. By tuning and optimizing this step, one can still improve the sparsity of the result.

We take the heart shape target excitation profile (fig-8.3-*top-left*).

In (fig-8.3-*top-right*) is the error decaying curve using the subspace sparse approximation method. The x axis is the number of the participating atoms. The y axis is the representation error (RRMS).

As expected, the representation error decays very fast as the number of the participating k sampling locations increases. Here is some representative data:

	TOL	Number of Atoms	Representation Error
A).	fixed 4× undersampl. in k_y	297	13.33%
B).	5%	67	4.90%
C).	10%	30	9.98%
D).	15%	18	14.67%
E).	20%	13	19.94%

Table 8.1: Subspace OMP error-decaying behavior

The excitation profiles and their according k space sampling trajectory regarding different accuracy requirement are shown in (fig 8.3).

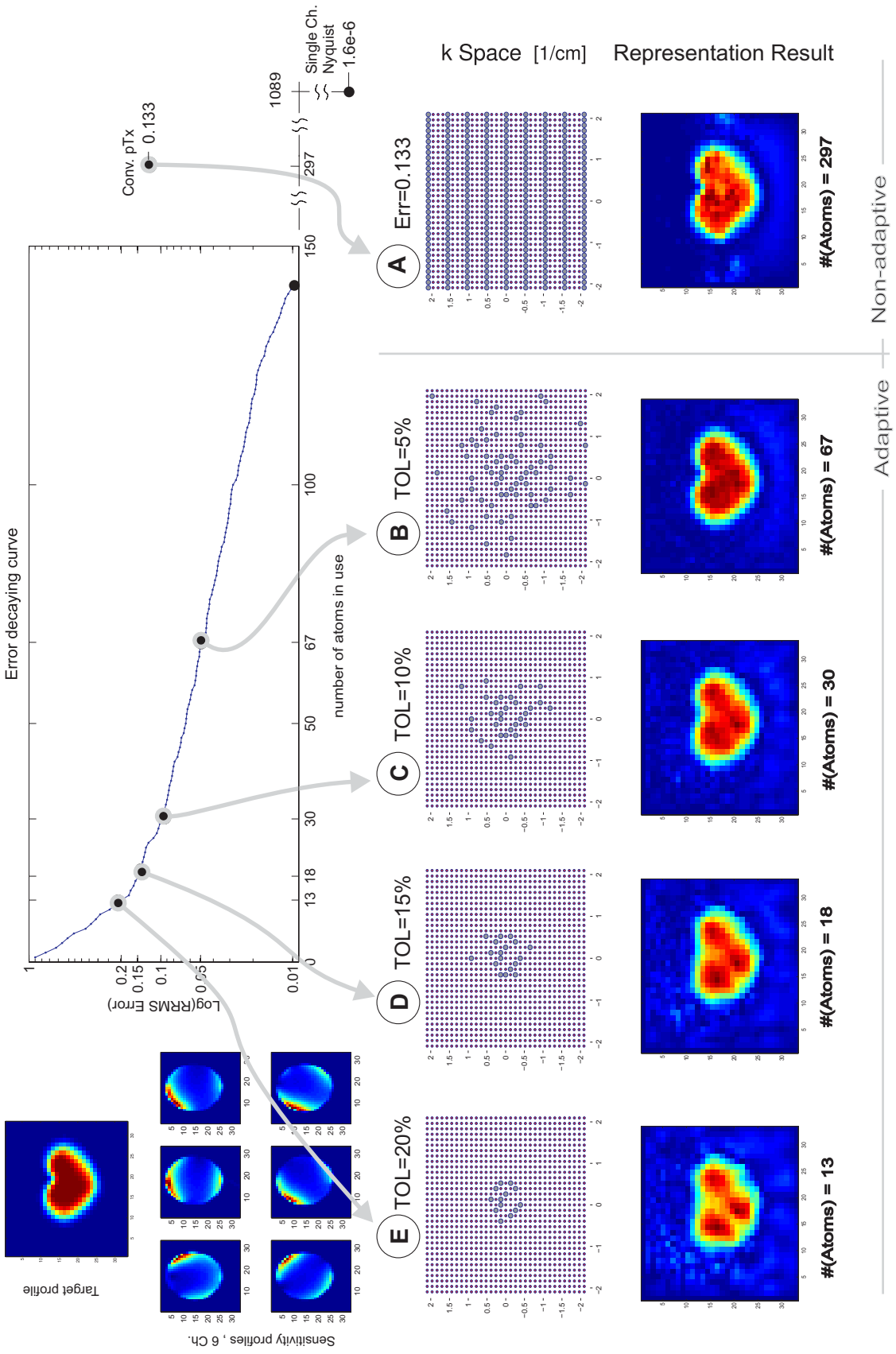


Figure 8.3: The validation of the subspace OMP algorithm for adaptive pTx

8.5 Validation Stage II: Bloch Equation Simulation

The targets of the stage II is are solving $\boldsymbol{P}_{\text{kTrav}}$ and $\boldsymbol{P}_{\text{RF-design}}$, validating the result via the Bloch equation simulation. The validation aim of this stage is the checklist No.2, *the transfer from sparsity to acceleration*, and the checklist No.3, *the availability of the STA assumption*.

As in the single channel setting, we go through the following steps: 1) $\boldsymbol{P}_{\text{kTrav}}$, 2) $\boldsymbol{P}_{\text{RF-design}}$, 3) $k(t) \mapsto g(t)$, 4) Bloch equation simulation with RF(t) and $g(t)$.

Unless explicitly mentioned, the most implementation details here are the same as in the single channel setting (§-7.7). One minor difference to the single channel validation is in approaching $\boldsymbol{P}_{\text{kTrav}}$: the global approach in (§-7.4) is applied here, while in the single channel setting we used the local approach.

(Fig-8.4) demonstrates the paradigm concretely for the pTx example case-D: (TOL=15%,$R = 13.9$).

The transfer from the sparsity to the actual acceleration is demonstrated in the plot of actual feasible sampling trajectory duration vs. the number of the obligatory sampling locations in k space (fig-8.5).

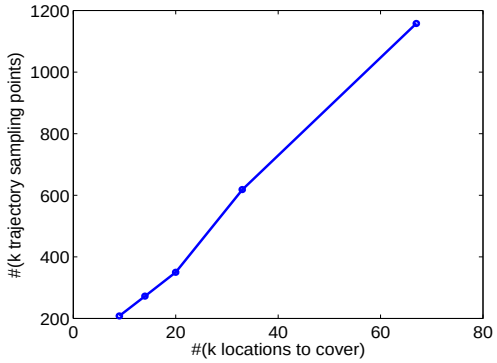


Figure 8.5: The sparsity vs. actual duration
x-axis: The number of selected k locations; *y-axis:* The k trajectory length (#(sampling-points) with $4mus$ sampling distance).

The linear relationship confirms that the sparsity can be sufficiently transferred to the actual acceleration.

As in the single channel case, since at the end more k space sampling points are used than the obligatory locations, the actual excitation quality

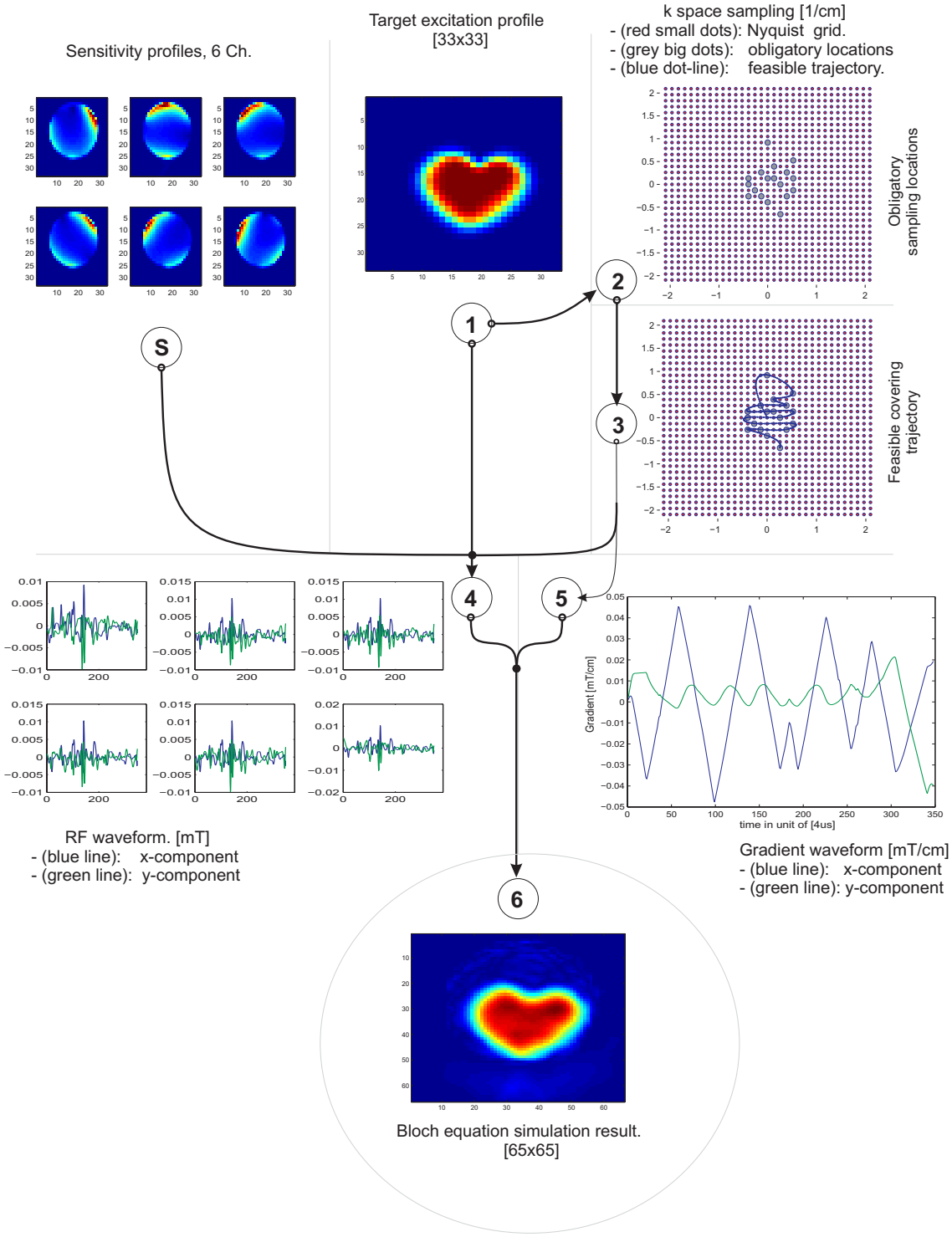


Figure 8.4: The example case-D, pTx setting

predicted by the Bloch equation simulation is usually better than the accuracy requirement TOL. (Fig-8.6) shows the correlation between the TOL and the actual error predicted by the simulation.

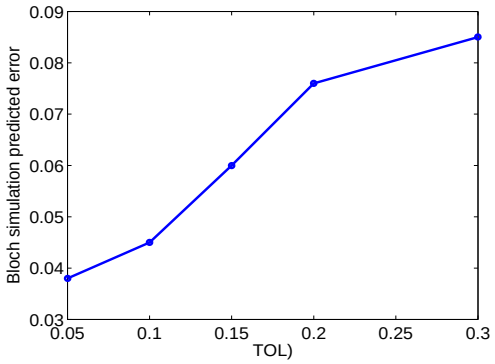


Figure 8.6: TOL vs. simulation predicted actual error

The plot of (fig-8.7-*top-right*) shows the actual error vs. excitation duration behavior.

In table-8.2 is some representative data. As in the single channel case, we use the fixed Nyquist full-sampling solution with 15.7ms excitation duration as the reference to measure the acceleration. The reduction factor $\boldsymbol{R}$ is defined as

$$\boldsymbol{R} := \frac{\text{Excitation duration [ms]}}{15.7\,\text{ms}}.$$

	TOL	Excitation Duration	Reduction Factor	Excitation Error
A).	4× undersamp.	3.68ms	$R = 4.3$	9.5%
B).	5%	2.47ms	$R = 6.35$	4.5%
C).	10%	1.4ms	$R = 11.2$	6.0%
D).	15%	1.13ms	$R = 13.9$	7.6%
E).	20%	0.83ms	$R = 18.9$	8.5%

Table 8.2: Some Bloch simulation data for pTx setting.

The excitation profiles and their according k space trajectories regarding different accuracy requirement are shown in the bottom blocks of (fig-8.7).

Compared with the conventional non adaptive parallel transmit paradigm, the adaptive approach can achieve better quality with stronger acceleration at the same time (fig-8.7: case-B vs case-A). Or with similar quality the adaptive approach can achieve much stronger acceleration, 18.9 vs. 4.3 (fig-8.7: case-E vs. case-A).

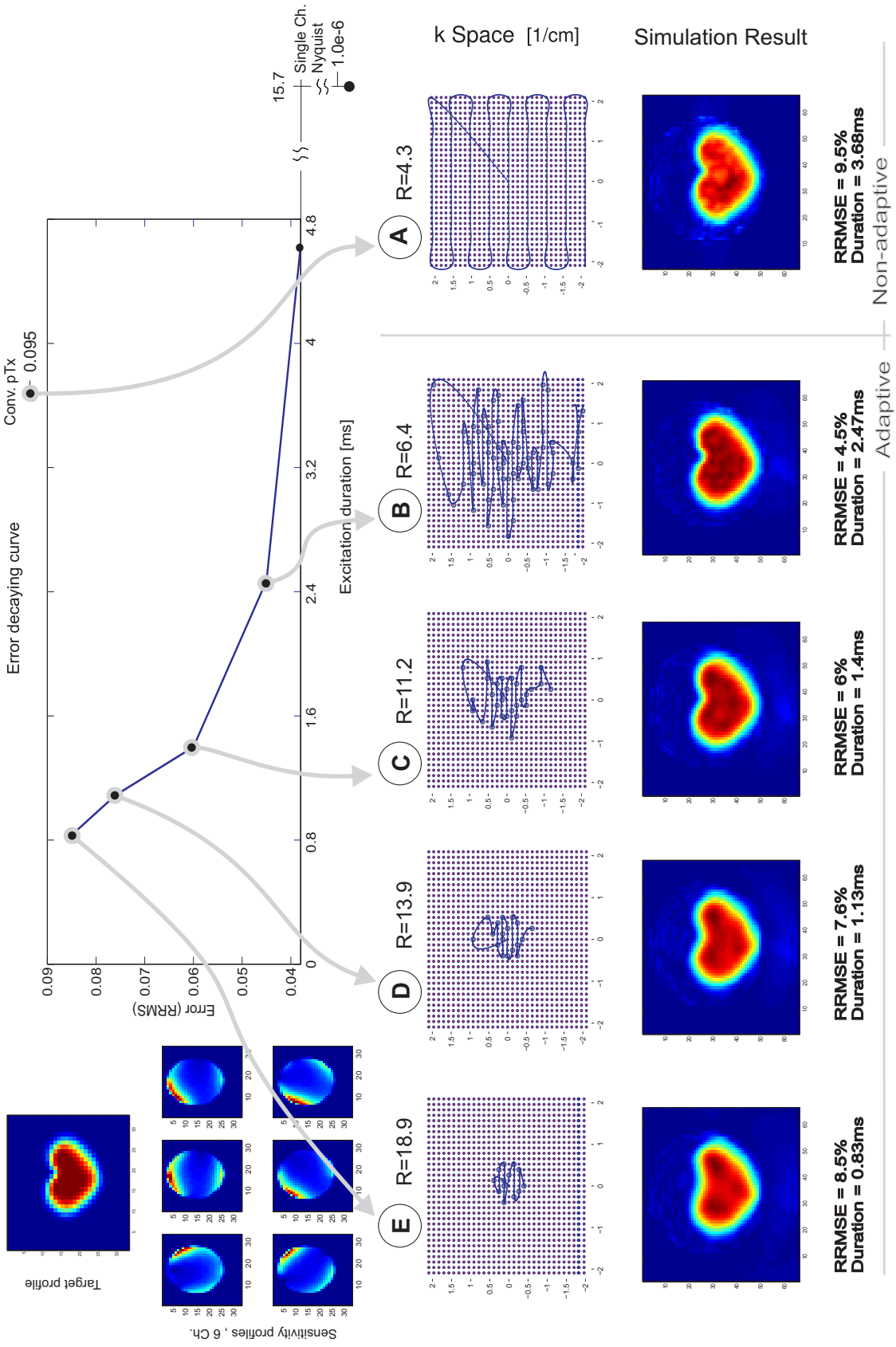


Figure 8.7: The validation results via the relaxation-free Bloch equation simulation, pTx setting

8.6 Validation Stage III: Phantom Experiments

The phantom experiments are performed on a GE MRI 3T scanner with dual channel Body coils. The same water phantom (ball-shape $d \sim 15cm$) as in the single channel case is used (fig-7.11).

Due to some hardware system issues, we performed the phantom experiment validation data by a two channels parallel transmit setting (instead of six or eight channels). But for the aim of proof of principle, it is sufficient to demonstrate the practical applicability of our concept.

(Fig-8.8-*top-left*) shows the target profile we used for the phantom experiment. (Fig-8.8-*top-right*) shows the RF transmit sensitivity profiles for the two RF channels.

The experiments are conducted at two difference reduction levels, $R = 2.0$ and $R = 2.8$. The k space sampling, Bloch equation simulation and the phantom images are shown in the middle and bottom rows in (fig-8.8). In the k space sampling plots, the big dots are the sparse obligatory sampling locations and the blue dot-line are the feasible covering trajectory.

The phantom images essentially reproduced the simulation results. The heart shape are good reproduced. In the reduction level $R = 2.8$ the phantom image also showed the same trend of increasing artifact as predicted by the simulation (at the left boundary area of the image). The matching of the simulated images and the phantom experiment images provided strong evidence for the praxis applicability of the proposed adaptive sparse pTx method.

Remark: Although the presented phantom experiments provide sufficient evidence for the first proof of principle, the quality of the phantom images are however less good as the ones for the single channel phantom experiment. Especially in both phantom images the left half of the heart shape are less excited than the right side, which are not expected. There could be two reasons for that:

1. the hardware synchronization of the two RF channels. This was also the issue preventing us from doing the 6 or 8 channel phantom experiment.

2. The quality of the sensitivity profiles.

Both of them, especially the accurate acquisition of the transmit sensitivity profile, are independent topics for themselves, which are beyond the topic

of this thesis. However, as one of the to-do items, further phantom experiments on more reliable hardware condition and sensitivity profiles should be performed to further quantitative analyze the practical applicability of the proposed adaptive parallel transmit.

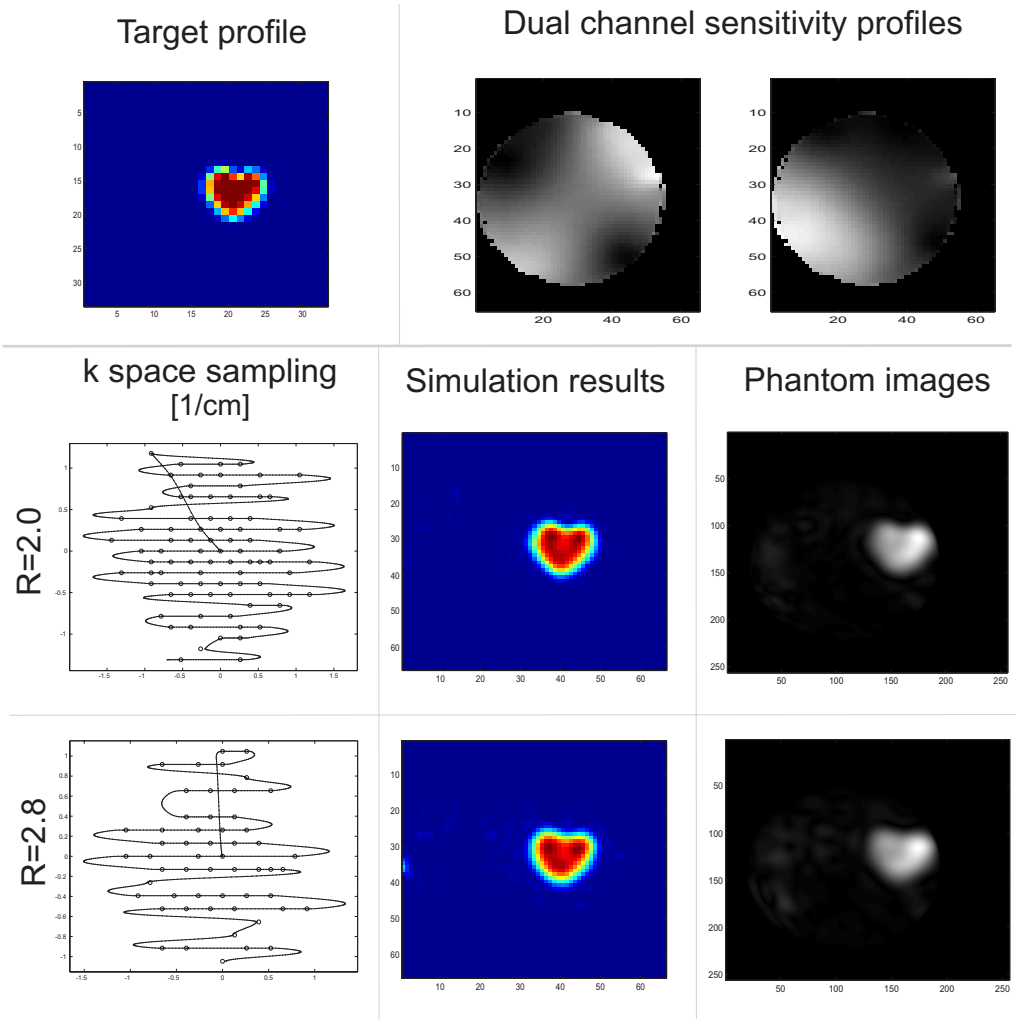


Figure 8.8: The validation results via phantom experiments, pTx setting

Chapter 9

Comparison and Conclusion

To conclude the advantage of using the adaptive paradigmen, for both single channel and pTx setting, we compare the four approaches (adaptive/non-adaptive, single-channel/pTx):

1. The conventional paradigma: single channel setting, non-adaptive k space sampling.

2. The non-adaptive pTx approach.

3. The adaptive approach with single transmit channel setting.

4. The adaptive approach with parallel transmit setting.

The comparison at the linear representation level is demonstrated in the (fig-9.1). The heart-shape target profile is used here. The error (y-axis) is plotted in logarithmic scalar.

The red dot-line is the error decaying curve for the adaptive approach with a six-channel parallel transmit setting. The blue dot-line is from the adaptive approach with the single channel setting. For example, for a user tolerance 10% the parallel transmit setting needs only about half so many obligatory locations to represent target profile as the single channel setting needs. This can be seen as the advatage of the dictionary overcompleteness provided by the parallel setting. From the sparse approximation point of view, it's an example that the overcompleteness of the dictinoary can indeed enhance the sparsity of the solution.

The solutions of the two non-adaptive approaches are represented with the green and black big dots for conventional pTx and conventional single

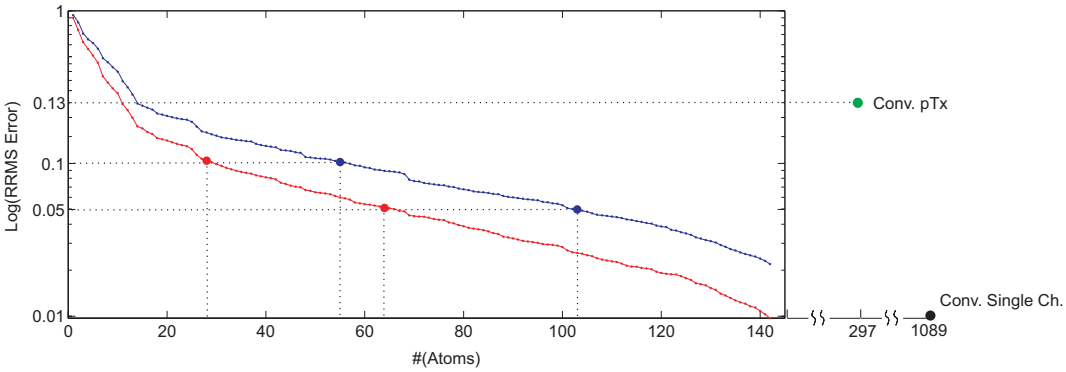


Figure 9.1: Comparisions of different methods: linear representation level. ((*red dot-line:*) adaptive pTx. (*blue dot-line:*) adaptive single ch. (*big green dot:*) non-adaptive pTx. (*big black dot:*) non-adaptive single ch.)

channel setting respectively. The advantage of the adaptive approach is very obvious.

The comparison at the level of actual excitation duration and excitatiaon quality is demonstrated with the (fig-9.2).

Comparing the adaptive with the non-adaptive approaches, the conclusion is quite obvious. For both single channel and pTx setting the adaptive paradigmen demonstrate strong acceleration capability. Taken the practical accuracy value $\sim 5\%$, the adaptive solution can provide very strong accerleration without any practically meaningful quality reduction. (We don't really care about the regime with error much smaller than 5%, since the hardware cannot realize the solution much better than 5%.)

Comparing the single channel adaptive approach with the parallel transmit adaptive approach, one also observes the advantage of the combination of the adaptivity and the parallelization. Especially for the low accuracy high reduction segment, the parallel transmit adaptive solution can clearly outperform the single channel adaptive solution. Compare case-III with case-IV or case-V, with similiar excitaiton quality the adaptive parallel transmit can further reduce the excitation duration to less than half of the adaptive single channel solution.

To conclude. In this work we've studied the acceleration of the spatial selective excitation process of magnetic resonance imaging. In contrast to the state-of-the-art non-adaptive approach we've developed a new paradigma based on adaptive concept. To our best knowledge this change from the

Shannon-Nyquist theorem based fixed sampling strategy to adaptive sampling location determination for acceleration is new in the community.

The adaptive concept is developed for both single channel setting and the emerging parallel transmit setting. With the adaptive parallel transmit we can combine the adaptive and the redundancy mechanism to maximize the acceleration capability.

The according mathematical tools are developed.

The proposed methods are implemented in Matlab. They are validated via both the Bloch equation simulation and the phantom experiments on a GE 3T scanner. The successful validation provides strong evidence for the practical applicability of the proposed methods.

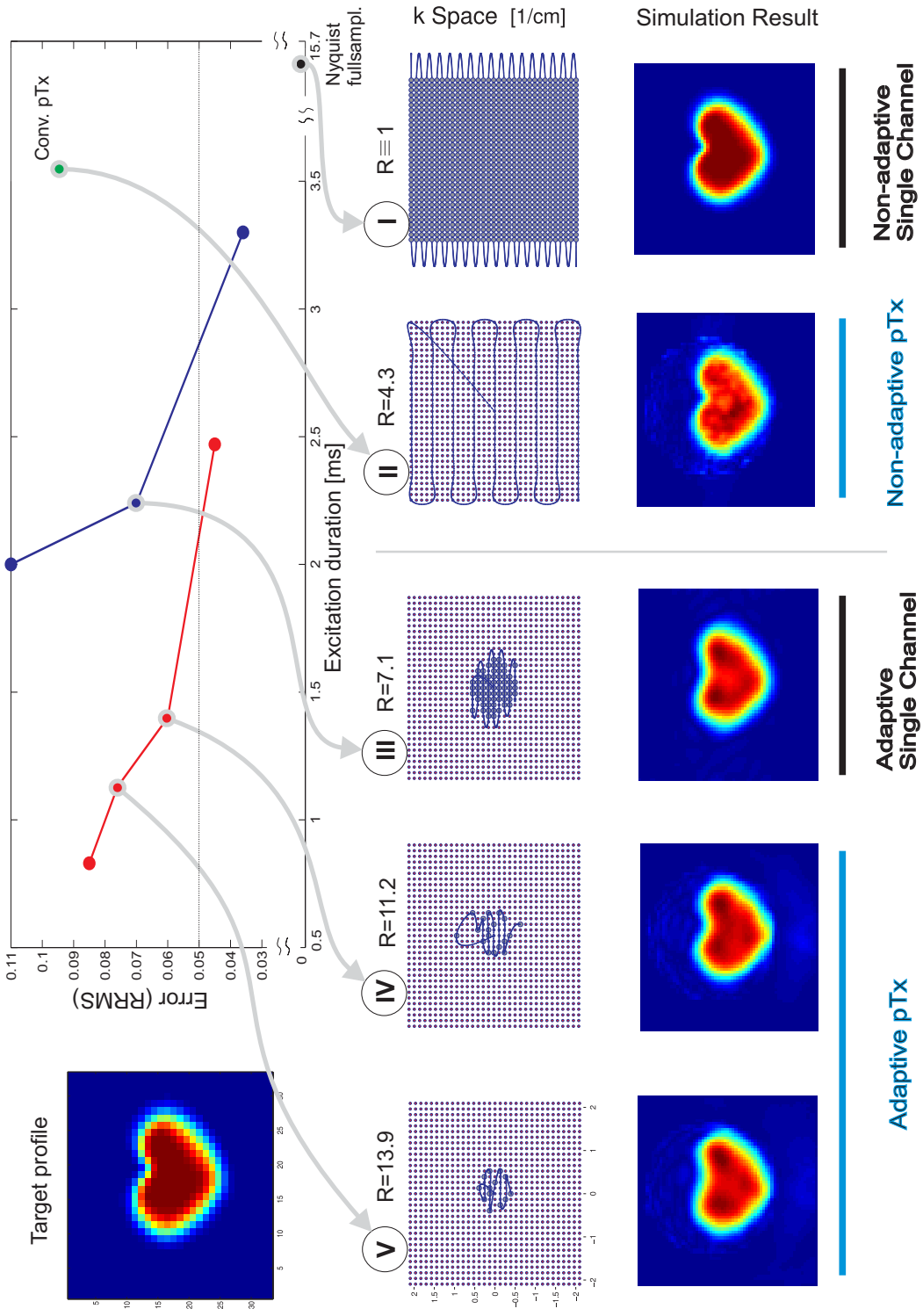


Figure 9.2: Comparisons of different methods: excitation duration vs. excitation quality

Bibliography

[BKZ04] Bernstein, King, and Zhou, *Handbook of mri pulse sequences*, Academic Press, 2004.

[Blo46] F Bloch, *Nuclear induction*, Phys. Rev. (1946).

[CDS99] S. Chen, D. Donoho, and M.A. Saunders, *Atomic decomposition by basis pursuit*, SIAM J. Sci. Comp. **20** (1999), 33–61.

[Che95] S. Chen, *Basis pursuit*, Ph.D. thesis, Statistics Dept., Stanford University, 1995.

[CNM86] S Conolly, D. Nishimura, and A. Macovski, *Optimal control solutions to the magnetic resonance selective excitation problem*, IEEE Trans Med Imaging (1986).

[CRT06] E.J. Candes, J. Romberg, and T. Tao, *Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information*, IEEE Transactions on Information Theory (2006).

[DB02] P. Deuflhard and F. Bornemann, *Scientific computing with ordinary differential equations*, Springer, 2002.

[DETR06] David L. Donoho, Michael Elad, Vladimir N. Temlyakov, and A. Overcomplete Representation, *Stable recovery of sparse overcomplete representations in the presence of noise*, IEEE Trans. Inform. Theory **52** (2006), 6–18.

[DH03] P. Deuflhard and A. Hohmann, *Numerical analysis in modern scientific computing*, Springer, 2003.

[DMA97] G Davis, S Mallat, and M Avellaneda, *Greedy adaptive approximation*, J. Constructive Approximation **13** (1997), 57–98.

[DMR+08] D. Donoho, A. Maleki, I. Rahman, M. Shahram, and V. Stodden, *15 years of reproducible research in computational harmonic analysis*, http://www.reproducibleresearch.org/DonohoEtAlCISESubmission.pdf (2008).

[Don06] D.L. Donoho, *Compressed sensing*, IEEE Transactions on Information Theory (2006).

[GB08] M. Grant and S. Boyd, *Lecture notes in control and information sciences*, Springer, 2008.

[GB09] ———, *Matlab software for disciplined convex programming (web page and software)*, http://stanford.edu/ boyd/cvx (2009).

[GBK08] U. Gamper, P. Boesiger, and S. Kozerke, *Compressed sensing in dynamic mri*, Magnetic Resonance in Medicine **59** (2008), 365373.

[GJH+02] M A Griswold, P M Jakob, R M Heidemann, M Nittka, V Jellus, J Wang, B Kiefer, and A Haase, *Generalized autocalibrating partially parallel acquisitions (grappa)*, Magn Reson Med (2002).

[GVL96] Gene H. Golub and Charles F. Van Loan, *Matrix computations (johns hopkins studies in mathematical sciences)*, The Johns Hopkins University Press, October 1996.

[GYZ+06] W Grissom, C Yip, Z Zhang, VA Stenger, JA Fessler, and D. Noll, *Spatial domain method for the design of rf pulses in multicoil parallel excitation*, Magn Reson Mad (2006).

[HBOR88] C. Hardy, P. Bottomley, M. ODonnell, and P. Roemer, *Optimization of two-dimensional spatially selective*, J Magn Reson (1988).

[HNC04] BA Hargreaves, DG Nishimura, and SM. Conolly, *Time-optimal multidimensional gradient waveform design for rapid imaging*, Magn Reson Med. **51** (2004), 81–92.

[Jay55] ET Jaynes, *Matrix treatment of nuclear induction*, Phys. Rev. (1955).

[JSN+09] H. Jung, K. Sung, K. S. Nayak, E. Y. Kim, and J. C. Ye, *k-t focuss: A general compressed sensing framework for high resolution dynamic mri*, Magnetic Resonance in Medicine **61** (2009), 103–116.

[KBLvdB03] U Katscher, P Börnert, C Leussler, and J S van den Brink, *Transmit sense*, Magn Reson Med (2003).

[Kin08] K. King, *Combining compressed sensing with parallel imaging (1488)*, ISMRM Toronto, 2008.

[LDP07] M Lustig, D Donoho, and JM Pauly, *Sparse mri: The application of compressed sensing for rapid mr imaging*, Magn Reson Med (2007).

[LDSP08] M Lustig, D Donoho, JM Santos, and JM Pauly, *Compressed sensing mri*, IEEE Signal Processing (2008).

[LKP08] M. Lustig, S-J Kim, and J. M. Pauly, *A fast method for designing time-optimal gradient waveforms for arbitrary k-space trajectories*, IEEE TMI **27** (2008), 866–873.

[LKS+09] Bo Liu, Kevin King, Michael Steckner, Jun Xie, Jinhua Sheng, and Leslie Ying, *Regularized sensitivity encoding (sense) reconstruction using bregman iterations*, Magnetic Resonance in Medicine **61** (2009), 145–152.

[LL99] Z. Liang and Lauterbur, *Principles of magnetic resonance imaging: A signal processing perspective*, Wiley-IEEE Press, 1999.

[Lus] M. Lustig, *timeoptimalgradient*, Stanford.

[MHB08] L. Marinelli, C. Hardy, and D. Blezek, *Mri with accelerated multi-coil compressed sensing (1484)*, ISMRM Toranto, ISMRM, 2008.

[MZ93] Stphane Mallat and Zhifeng Zhang, *Matching pursuit with time-frequency dictionaries*, IEEE Transactions on Signal Processing **41** (1993), 3397–3415.

[Nat95] B. K. Natarajan, *Sparse approximate solutions to linear systems*, SIAM Journal on Computing **24** (1995), no. 2, 227–234.

[Pat93] *Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition*, 1993.

[PNM89a] J. Pauly, D. Nishimura, and A. Macovski, *A k-space analysis of small-tip-angle excitation*, J Magn Reson (1989).

[PNM89b] J M Pauly, D G Nishimura, and A Macovski, *A linear class of large-tip-angle selective excitation pulses*, J Magn Reson **82** (1989), 571587.

[Pru04] K P Pruessmann, *Parallel imaging at high field strength: Synergies and joint potential*, Topics in Magetic Resonance Imaging (2004).

[PS82] Christopher C. Paige and Michael A. Saunders, *Lsqr: An algorithm for sparse linear equations and sparse least squares*, ACM Trans. Math. Softw. **8** (1982), no. 1, 43–71.

[PWSB99] K P Pruessmann, M Weiger, M B Scheidegger, and P Boesiger, *Sense: Sensitivity encoding for fast mri*, Magn Reson Med (1999).

[SS90] G. W. Stewart and Ji-Guang Sun, *Matrix perturbation theory (computer science and scientific computing) (computer science and scientific computing)*, Academic Press, June 1990.

[TGS06] Joel A. Tropp, Anna C. Gilbert, and Martin J. Strauss, *Algorithms for simultaneous sparse approximation. part i: Greedy pursuit*, Signal Processing **86** (2006), no. 3, 572 – 588, Sparse Approximations in Signal and Image Processing.

[Tro04a] J. Tropp, *Topics in sparse approximation*, Ph.D. thesis, University of Texas at Austin, 2004.

[Tro04b] J.A Tropp, *Greed is good: algorithmic results for sparse approximation*, IEEE Transactions on Information Theory (2004), 2231 – 2242.

[Tro06] Joel A. Tropp, *Algorithms for simultaneous sparse approximation. part ii: Convex relaxation*, Signal Processing **86** (2006), no. 3, 589 – 602, Sparse Approximations in Signal and Image Processing.

[XKZ+08] D. Xu, K. King, Y. Zhu, G. McKinnon, and Z. Liang, *Designing multichannel, multidimensional, arbitrary flip angle rf pulses using an optimal control approach*, Magn Reson Med (2008).

[Zhu04] Y Zhu, *Parallel excitation with an array of transmit coils*, Magn Reson Med (2004).

[ZWS+07] A. C. Zelinski, L. L. Wald, K. Setsompop, V. Alagappan, B. A. Gagoski, V. K Goyal, F. Hebrank, U. Fontius, F. Schmitt, and E. Adalsteinsson, *Comparison of 3 algorithms for solving linearized systems of parallel excitation rf waveform design equations: Experiments on an 8-channel system at 3t*, Concepts in Magnetic Resonance, Part B: Magnetic Resonance Engineering **31B** (2007).