EE290T: Advanced Reconstruction Methods for Magnetic Resonance Imaging

Martin Uecker
Topics:

- Image Reconstruction as Inverse Problem
- Parallel Imaging
- Non-Cartesian MRI
- Nonlinear Inverse Reconstruction
- k-space Methods
- Subspace Methods
- Model-based Reconstruction
- Compressed Sensing
Tentative Syllabus

- 01: Jan 27 Introduction
- 02: Feb 03 Parallel Imaging as Inverse Problem
- 03: Feb 10 Iterative Reconstruction Algorithms
- —: Feb 17 (holiday)
- 04: Feb 24 Non-Cartesian MRI
- 05: Mar 03 Nonlinear Inverse Reconstruction
- 06: Mar 10 Reconstruction in k-space
- 07: Mar 17 Reconstruction in k-space
- —: Mar 24 (spring recess)
- 08: Mar 31 Subspace methods
- 09: Apr 07 Model-based Reconstruction
- 10: Apr 14 Compressed Sensing
- 11: Apr 21 Compressed Sensing
- 12: Apr 28 TBA
Projects

- Two pre-defined homework projects
- Final project
- Groups: 2-3 people

- 1. Project: iterative SENSE (Feb 03 - Feb 24)
- 2. Project: non-Cartesian MRI (Feb 24 - Mar 17)
- 3. Final Project (Mai 17 - Apr 28)

- Grading: 25%, 25%, 50%
Today

- Introduction to MRI
- Basic image reconstruction
- Inverse problems
Introduction to MRI

- Non-invasive technique to look “inside the body”
- No ionizing radiation or radioactive materials
- High spatial resolution
- Multiple contrasts
- Volumes or arbitrary crosssectional slices
- Many applications in radiology/neuroscience/clinical research
Disadvantages

Some disadvantages:

- Expensive
- **Long measurement times**
- Not every patient is admissible
  (cardiac pacemaker, implants, tattoo, ...)
MRI Scanner

- Magnet
- Radiofrequency coils (and electronics)
- Gradient system
- Superconducting magnet
- Field strength: 0.25 T - 9.4 T (human)
MRI Scanner

- Magnet
- Radiofrequency coils (and electronics)
- Gradient system
- Spin excitation
- Signal detection
MRI Scanner

- Magnet
- Radiofrequency coils (and electronics)
- Gradient system

- Creates magnetic field gradients on top of the static $B_0$ field
- Switchable gradients in all directions
Protons in the Magnetic Field

- Protons have spin 1/2 ("little magnets")
- Two energy levels $\Delta E = 2\hbar \gamma \frac{1}{2} B^0$
- Larmor precession $\Delta E = \hbar \omega_{\text{Larmor}}$
- Excitation with resonant radiofrequency pulses
- Macroscopic behaviour described by Bloch equations
Bloch Equations

Macroscopic (classical) behaviour of the magnetization corresponds to expectation value $\langle \vec{s} \rangle$ of spin operator (Pauli matrices).

Dynamics described by differential equations:

$$\frac{d}{dt} \langle \vec{s} \rangle = -\frac{e}{m_0} \langle \vec{s} \rangle \times \begin{pmatrix} B_x^1(t) \\ B_y^1(t) \\ B_z^0 \end{pmatrix} + \begin{pmatrix} -\frac{1}{T_2} \langle s_x \rangle \\ -\frac{1}{T_2} \langle s_y \rangle \\ \frac{s_0 - \langle s_z \rangle}{T_1} \end{pmatrix}$$
The Basic NMR-Experiment

Equilibrium

RF-Signal

Bloch equations:

\[
\frac{d}{dt} \langle \vec{s} \rangle = -\frac{e}{m_0} \langle \vec{s} \rangle \times \begin{pmatrix} B_x^1(t) \\ B_y^1(t) \\ B_z^0 \end{pmatrix} + \begin{pmatrix} -\frac{1}{T_2} \langle s_x \rangle \\ -\frac{1}{T_2} \langle s_y \rangle \\ \frac{s_0 - \langle s_z \rangle}{T_1} \end{pmatrix}
\]
The Basic NMR-Experiment

Bloch equations:

$$\frac{d}{dt} \langle \vec{s} \rangle = -\frac{e}{m_0} \langle \vec{s} \rangle \times \begin{pmatrix} B_x^1(t) \\ B_y^1(t) \\ B_z^0 \end{pmatrix} + \begin{pmatrix} -\frac{1}{T_2} \langle s_x \rangle \\ -\frac{1}{T_2} \langle s_y \rangle \\ \frac{s_0 - \langle s_z \rangle}{T_1} \end{pmatrix}$$
The Basic NMR-Experiment

Precession

RF-Signal

Bloch equations:

\[
\frac{d}{dt} \langle \vec{s} \rangle = -\frac{e}{m_0} \langle \vec{s} \rangle \times \begin{pmatrix} B_x^1(t) \\ B_y^1(t) \\ B_z^0 \end{pmatrix} + \begin{pmatrix} -\frac{1}{T_2} \langle s_x \rangle \\ -\frac{1}{T_2} \langle s_y \rangle \\ \frac{s_0 - \langle s_z \rangle}{T_1} \end{pmatrix}
\]
The Basic NMR-Experiment

$T_2$-decay

Bloch equations:

$$\frac{d}{dt} \langle \vec{s} \rangle = -\frac{e}{m_0} \langle \vec{s} \rangle \times \begin{pmatrix} B_x^1(t) \\ B_y^1(t) \\ B_z^0 \end{pmatrix} + \begin{pmatrix} -\frac{1}{T_2} \langle s_x \rangle \\ -\frac{1}{T_2} \langle s_y \rangle \\ \frac{s_0 - \langle s_z \rangle}{T_1} \end{pmatrix}$$
The Basic NMR-Experiment

Bloch equations:

\[
\frac{d}{dt} \langle \vec{s} \rangle = -\frac{e}{m_0} \langle \vec{s} \rangle \times \begin{pmatrix} B_x^1(t) \\ B_y^1(t) \\ B_z^0 \end{pmatrix} + \begin{pmatrix} -\frac{1}{T_2} \langle s_x \rangle \\ -\frac{1}{T_2} \langle s_y \rangle \\ s_0 - \frac{\langle s_z \rangle}{T_1} \end{pmatrix}
\]
Multiple Contrasts

- Measurement of various physical quantities possible
- Advantage over CT / X-Ray (measurement of tissue density)

Proton density | T1-weighted | T2-weighted
Receive Chain

Thermal noise:
- Sample noise
- Coil noise
- Preamplifier noise

phased-array coil
$R = R_C + R_S$

- $R_C$ loss in coil
- $R_S$ loss in sample

$\Rightarrow$ Johnson-Nyquist noise $v_n$
Mean-square noise voltage:

\[ \overline{v_n^2} = 4k_B T R BW \]

- additive Gaussian white noise
- Brownian motion of electrons in body and coil
- Connection between resistance and noise: Fluctuation-Dissipation-Theorem
Spatial Encoding

Nobel price in medicine 2003 for Paul C. Lauterbur and Sir Peter Mansfield

Principle:
- Additional field gradients
  ⇒ Larmor frequency dependent on location

Concepts:
- Slice selection (during excitation)
- Phase/Frequency encoding (before/during readout)

PC Lauterbur, Image formation by induced local interactions: examples employing nuclear magnetic resonance, Nature, 1973
Frequency Encoding

- Additional field gradients $B^0(x) = B^0 + G_x \cdot x$
  - Frequency $\omega(x) = \gamma B^0(x)$

\[\begin{align*}
\vec{k}(t) &= \gamma \int_0^t d\tau \, \vec{G}(\tau) \\
\mathbf{s}(t) &= \int_V d\vec{x} \, \rho(\vec{x}) e^{-i2\pi \vec{k}(t) \cdot \vec{x}}
\end{align*}\]

(ignoring relaxation and errors)
Frequency Encoding

- Additional field gradients $B^0(x) = B^0 + G_x \cdot x$

  $\Rightarrow$ Frequency $\omega(x) = \gamma B^0(x)$

\[
\begin{align*}
\vec{k}(t) &= \gamma \int_0^t d\tau \vec{G}(\tau) \\
s(t) &= \int_V d\vec{x} \rho(\vec{x}) e^{-i2\pi \vec{k}(t) \cdot \vec{x}}
\end{align*}
\]

(ignoring relaxation and errors)
Gradient System

- z gradient
- y gradient
- x gradient
Gradient System

- z gradient
- y gradient
- x gradient
Gradient System

- z gradient
- y gradient
- x gradient
Gradient System

- z gradient
- y gradient
- x gradient
Gradient System

- z gradient
- y gradient
- x gradient
Imaging Sequence

FLASH (Fast Low Angle SHot)
Imaging Sequence

readout (in-plane)

radio

slice

read

phase

repetition time TR

FLASH (Fast Low Angle SHot)
Fourier Encoding

radio slice read phase

repeation time TR

$k_{phase}$

$k_{read}$
Fourier Encoding

radio
slice
read
phase

\[ \alpha \]

acquisition

repetition time TR

\[ k_{phase} \]

\[ k_{read} \]
Image Reconstruction

k-space data
Image Reconstruction

inverse Discrete Fourier Transform
Direct Image Reconstruction

- **Assumption:** Signal is Fourier transform of the image:

\[
s(t) = \int d\vec{x} \, \rho(\vec{x}) e^{-i2\pi \vec{x} \cdot \vec{k}(t)}
\]

- Image reconstruction with inverse DFT

\[
\vec{k}(t) = \gamma \int_0^t d\tau \, \tilde{G}(\tau)
\]

⇒ sampling

⇒ k-space

⇒ iDFT

⇒ image
Requirements

▶ Short readout (signal equation holds for short time spans only)
▶ Sampling on a Cartesian grid
⇒ Line-by-line scanning

measurement time:
TR ≥ 2ms
2D: $N \approx 256 \Rightarrow$ seconds
3D: $N \approx 256 \times 256 \Rightarrow$ minutes
Sampling
Sampling
Poisson Summation Formula

Periodic summation of a signal $f$ from discrete samples of its Fourier transform $\hat{f}$:

$$\sum_{n=-\infty}^{\infty} f(t + nP) = \sum_{l=-\infty}^{\infty} \frac{1}{P} \hat{f} \left( \frac{l}{P} \right) e^{2\pi i \frac{l}{P} t}$$

No aliasing $\Rightarrow$ Nyquist criterion

For MRI: $P = \text{FOV} \Rightarrow k$-space samples are at $k = \frac{l}{\text{FOV}}$
k-space

periodic summation from discrete Fourier samples on grid

other positions can be recovered by sinc interpolation (Whittaker-Shannon formula)
- periodic summation from discrete Fourier samples on grid
- periodic summation from discrete Fourier samples on grid
- other positions can be recovered by sinc interpolation (Whittaker-Shannon formula)
Nyquist-Shannon Sampling Theorem

**Theorem 1:** If a function $f(t)$ contains no frequencies higher than $W$ cps, it is completely determined by giving its ordinates at a series of points spaced $1/2W$ seconds apart.\(^1\)

- Band-limited function
- Regular sampling
- Linear sinc-interpolation

---

Sampling
Point Spread Function

- **Signal:**

  \[ y = P \mathcal{F} x \]

  sampling operator \( P \), Fourier transform \( \mathcal{F} \)

- **Reconstruction with (continuous) inverse Fourier transform:**

  \[ \mathcal{F}^{-1} P^H P \mathcal{F} x = psf \star x \]

  \( \Rightarrow \) Convolution with point spread function (PSF)

  (shift-invariance)

- **Question:** What is the PSF for the grid?
Dirichlet Kernel

\[ D_n(x) = \sum_{k=-n}^{n} e^{2\pi ikx} = \frac{\sin(\pi(2n+1)x)}{\sin(\pi x)} \]

\[(D_n \ast f)(x) = \int_{-\infty}^{\infty} dy \ f(y) D_n(x - y) = \sum_{k=-n}^{n} \hat{f}(k) e^{2\pi ikx} \]

FWHM \approx 1.2

\[ D_{16} \]
Gibbs Phenomenon

- Truncation of Fourier series
- \(\Rightarrow\) **Ringing** at jump discontinuities

Rectangular wave and Fourier approximation
Gibbs Phenomenon

- Truncation of Fourier series
  ⇒ Ringing at jump discontinuities

Rectangular wave and Fourier approximation
Gibbs Phenomenon

- Truncation of Fourier series
  - Ringing at jump discontinuities

Rectangular wave and Fourier approximation
Gibbs Phenomenon

- Truncation of Fourier series

⇒ **Ringing** at jump discontinuities

Rectangular wave and Fourier approximation
Assumptions:

- Signal is Fourier transform of the image
- Sampling on a Cartesian grid
- Signal from a limited (compact) field of view
- Missing high-frequency samples are small
- Noise is neglectable
Fast Fourier Transform (FFT) Algorithms

Discrete Fourier Transform (DFT):

\[
DFT_k^n \{f_n\}_{n=0}^{N-1} := \hat{f}_k = \sum_{n=0}^{N-1} e^{i \frac{2\pi}{N} kn} f_n \quad \text{for} \quad k = 0, \ldots, N - 1
\]

Matrix multiplication: \(O(N^2)\)

Fast Fourier Transform Algorithms: \(O(N \log N)\)
Cooley-Tukey Algorithm

Decomposition:

\[ N = N_1 N_2 \]
\[ k = k_2 + k_1 N_2 \]
\[ n = n_2 N_1 + n_1 \]

With \( \xi_N := e^{i \frac{2\pi}{N}} \) we can write:

\[
(\xi_N)^{kn} = (\xi_{N_1 N_2})^{(k_2 + k_1 N_2)(n_2 N_1 + n_1)} \\
= (\xi_{N_1 N_2})^{N_2 k_1 n_1} (\xi_{N_1 N_2})^{k_2 n_1} (\xi_{N_1 N_2})^{N_1 k_2 n_2} (\xi_{N_1 N_2})^{N_1 N_2 k_1 n_2} \\
= (\xi_{N_1})^{k_1 n_1} (\xi_{N_1 N_2})^{k_2 n_1} (\xi_{N_2})^{k_2 n_2}
\]

(use: \( \xi_{AB} = \xi_B \) and \( \xi_A = 1 \))
Cooley-Tukey Algorithm

Decomposition of a DFT:

\[
\begin{align*}
DFT_k^N \{ f_n \}_{n=0}^{N-1} &= \sum_{n=0}^{N-1} (\xi_N)^{kn} f_n \\
&= \sum_{n_1=0}^{N_1-1} (\xi_{N_1})^{k_1 n_1} (\xi_N)^{k_2 n_1} \sum_{n_2=0}^{N_2-1} (\xi_{N_2})^{k_2 n_2} f_{n_1+n_2 N_1} \\
&= DFT_{k_1}^{N_1} \left\{ (\xi_N)^{k_2 n_1} DFT_{k_2}^{N_2} \{ f_{n_1+n_2 N_1} \}_{n_2=0}^{N_2-1} \right\}_{n_1=0}^{N_1-1}
\end{align*}
\]

- Recursive decomposition to prime-sized DFTs
- Efficient computation for smooth sizes

( efficient algorithms available for all other sizes too)
Definition (Hadamard): A problem is called well-posed if

1. there exists a solution to the problem (existence),
2. there is at most one solution to the problem (uniqueness),
3. the solution depends continuously on the data (stability).

(we will later see that all three conditions can be violated)
Moore-Penrose Generalized Inverse

(A linear and bounded)

\( x \) is least-squares solution, if

\[
\|Ax - y\| = \inf \{\|A\hat{x} - y\| : \hat{x} \in X\}
\]

\( x \) is best approximate solution, if

\[
\|x\| = \inf \{\|\hat{x}\| : \hat{x} \text{ least-squares solution}\}
\]

Generalized inverse \( A^\dagger : D(A^\dagger) := R(A) + R(A)^\perp \mapsto X \) maps data to the best approximate solution.
Tikhonov Regularization

\[(\text{inverse not continuous: } \|A^{-1}\| = \infty)\]

Regularized optimization problem:

\[x_\delta^\alpha = \arg\min_x \|Ax - y^\delta\|_2^2 + \alpha \|x\|_2^2\]

Explicit solution:

\[x_\delta^\alpha = \left(\left(A^H A + \alpha I\right)^{-1} A^H y^\delta\right)_{A_\delta^\alpha}^{A^\dagger}\]

Generalized inverse:

\[A^\dagger = \lim_{\alpha \rightarrow 0} A_\delta^\alpha\]
Tikhonov Regularization: Bias vs Noise

Noise contamination:

\[ y^\delta = y + \delta n \]
\[ = Ax + \delta n \]

Reconstruction error:

\[ x - x_\alpha^\delta = x - (A^H A + \alpha I)^{-1} A^H y^\delta \]
\[ = x - (A^H A + \alpha I)^{-1} A^H Ax + (A^H A + \alpha I)^{-1} A^H \delta n \]
\[ = \alpha (A^H A + \alpha I)^{-1} x + (A^H A + \alpha I)^{-1} A^H \delta n \]

\[ \text{approximation error} \quad \text{data noise error} \]

Morozov’s discrepancy principle:

Largest regularization with \[ \| Ax_\alpha^\delta - y^\delta \|_2 \leq \tau \delta \]
Example

\[ y = \begin{pmatrix} 3.4525 & 3.3209 & 3.3090 \\ 3.3209 & 3.3706 & 3.3188 \\ 3.3090 & 3.3188 & 3.2780 \end{pmatrix} \begin{pmatrix} 1 \\ 0.5 \\ 0.2 \end{pmatrix} + \begin{pmatrix} 1.6522 \times 10^{-2} \\ 1.5197 \times 10^{-3} \\ 3.3951 \times 10^{-3} \end{pmatrix} = \begin{pmatrix} 5.7923 \\ 5.6717 \\ 5.6274 \end{pmatrix} \]

Ill-conditioning: \( \text{cond}(A) = \frac{\sigma_{\text{max}}}{\sigma_{\text{min}}} \approx 10000 \)

Solution:
(\( \alpha = 0 \))
\[ \hat{x} = A^{-1}y = \begin{pmatrix} 1.25414 \\ 1.02660 \\ -0.58866 \end{pmatrix} \]

Regularized solution:
(\( \alpha = 3 \times 10^{-6} \))
\[ \hat{x}_{\alpha} = A_{\alpha}^\dagger y = \begin{pmatrix} 1.09706 \\ 0.45751 \\ 0.14632 \end{pmatrix} \]
Tikhonov Regularization using SVD

Singular Value Decomposition (SVD):

\[ A = U \Sigma V^H = \sum_j \sigma_j U_j V_j^H \]

Regularized inverse:

\[ A_\alpha^\dagger = \left( A^H A + \alpha I \right)^{-1} A^H = \sum_j \frac{\sigma_j}{\sigma_j^2 + \alpha} V_j U_j^H \]

Approximation error:

\[ 1 - \frac{\sigma_j^2}{\sigma_j^2 + \alpha} = \frac{\alpha}{\sigma_j^2 + \alpha} \]

Data noise error:

\[ \frac{\sigma_j \delta}{\sigma_j^2 + \alpha} \]
Image Reconstruction as Inverse Problem

Forward problem:

\[ y = Ax + n \]

\( x \) image (and more), \( A \) forward operator, \( n \) noise, \( y \) data

Regularized solution:

\[ x^* = \text{argmin}_x \left( \|Ax - y\|^2_2 + \alpha R(x) \right) \]

\( \alpha \) data consistency + regularization

Advantages:

- Simple extension to non-Cartesian trajectories
- Modelling of physical effects
- Prior knowledge via suitable regularization terms
Extension to Non-Cartesian Trajectories

Practical implementation issues:

- Imperfect gradient waveforms (e.g. delays, eddy currents)
- Efficient implementation of the reconstruction algorithm
Modelling of Physical Effects

Examples:

- Coil sensitivities (parallel imaging)
  \[ s_j(t) = \int d\vec{x} \, \rho(\vec{x}) c_j(\vec{x}) e^{-i2\pi \vec{x} \cdot \vec{k}(t)} \]

- $T_2$ relaxation
  \[ s(t) = \int d\vec{x} \, \rho(\vec{x}) e^{-R(\vec{x}) t} e^{-i2\pi \vec{x} \cdot \vec{k}(t)} \]

- Field inhomogeneities
  \[ s(t) = \int d\vec{x} \, \rho(\vec{x}) e^{-i\Delta B_0(\vec{x}) t} e^{-i2\pi \vec{x} \cdot \vec{k}(t)} \]

- Diffusion, flow, motion, ...
Regularization

- Introduces additional information about the solution
- In case of ill-conditioning: needed for stabilization

**Common choices:**

- **Tikhonov (small norm)**
  \[ R(x) = \| W(x - x_R) \|_2^2 \quad \text{(often: } W = I \text{ and } x_R = 0) \]

- **Total variation (piece-wise constant images)**
  \[ R(x) = \int d\vec{r} \sqrt{\left| \partial_1 x(\vec{r}) \right|^2 + \left| \partial_2 x(\vec{r}) \right|^2} \]

- **\( L_1 \) regularization (sparsity)**
  \[ R(x) = \| W(x - x_R) \|_1 \]
Probability of $x$ given $y$: $p(x|y)$

Likelihood: Probability of a specific outcome given a parameter: $p(y|x)$

Maximum-Likelihood-Estimator: The estimate is the parameter $x$ which maximizes the likelihood.
Statistical Model

**Linear measurements** contaminated by noise:

\[ y = Ax + n \]

**Gaussian white noise:**

\[ p(n) = \mathcal{N}(0, \sigma^2) \text{ with } \mathcal{N}(\mu, \sigma^2) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]

Probability of an outcome (measurement) given the image \( x \):

\[ p(y|A, x, \lambda) = \mathcal{N}(Ax, \sigma^2) \]
Prior Knowledge

Prior knowledge as a probability distribution for the image:

\[ p(x) = \cdots \]

A posterior probability distribution \( p(x|y) \) given the data \( y \) can be computed using Bayes’ theorem:

\[ p(x|y) = \frac{p(y|x)p(x)}{p(y)} \]

A point estimate for the image can then be obtained, for example by maximum a posteriori estimation (MAP), or by minimizing the posterior expected value of a loss function, e.g. a minimum mean squared error (MMSE) estimator.
Bayesian Prior

- **$L_2$-Regularization: Ridge Regression**

\[ N(\mu, \sigma^2) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{Gaussian prior} \]

- **$L_1$-Regularization: LASSO**

\[ p(x|\mu, b) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}} \quad \text{Laplacian prior} \]

- **$L_2$ and $L_1$: Elastic net\(^1\)**

---

Summary

- Magnetic Resonance Imaging
- Direct image reconstruction
- Inverse problems