Diagnostic Medical Image Processing

MR Intensity Inhomogeneities

WS 2010/11

Joachim Hornegger, Dietrich Paulus, Markus Kowarschik

Lehrstuhl für Mustererkennung (Informatik 5)

Friedrich-Alexander-Universität Erlangen-Nürnberg

Diagnostic Medical Image Processing

MR Acquisition Devices - Revisited

MR scanners are huge and heavy systems with strong superconducting magnets:

Figure: The most recent 3 Tesla systems: Magnetom Verio and Trio Scanner (images: Siemens Medical Solutions)

Physical and Mathematical Ingredients

For MRI imaging the following physical and mathematical concepts are required to be understood:

- **nuclei** serve as objects to be imaged
- **homogeneous magnetic fields** are generated by the scanner to align the nuclear moment vectors
- **resonance phenomenon** that results from the interaction of nuclei with the magnetic field
- **Fourier methods** for image reconstruction
- **image enhancement algorithms** to compensate for violations of the required homogeneity of the magnetic field (bias field correction)

Note: Details in physics are not in the focus of our lecture, but the algorithmic aspects. We begin with **bias field correction**.

Components of an MR Scanner

The major four components of an MR Scanner are:

- main magnet
- magnetic field gradient system
- radio frequency system (RF system)
- imaging system

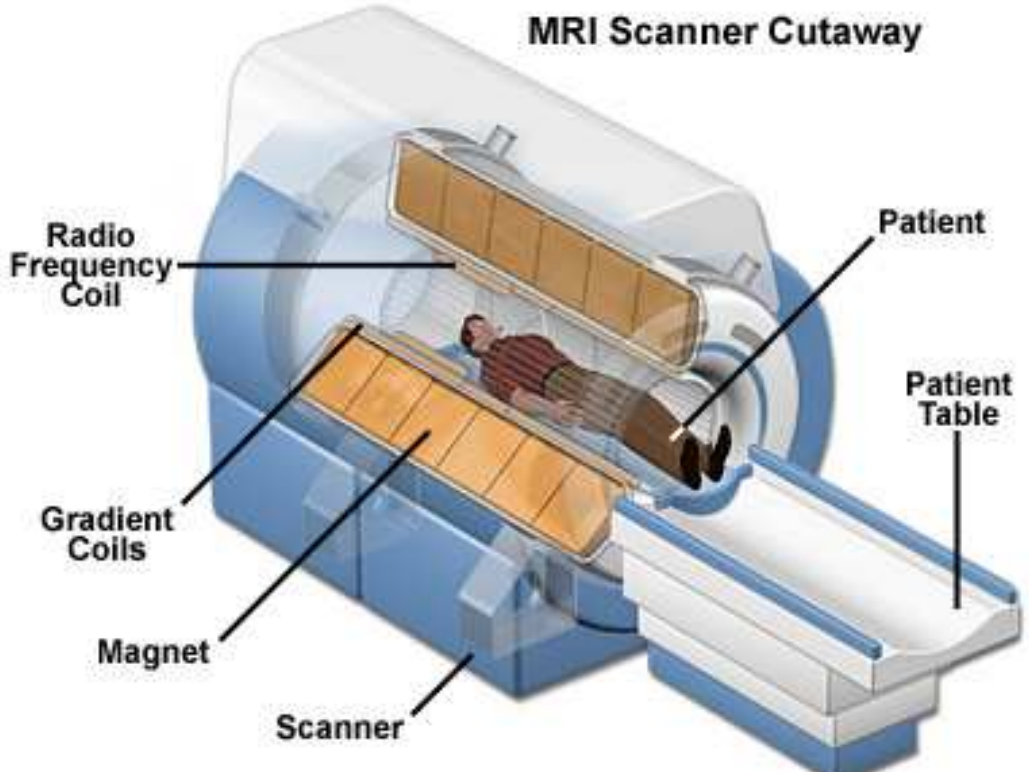


Figure: Main Components of an MR Scanner (image borrowed from here)

Components of an MR Scanner: Main Magnet

The **main magnet** is required to generate a strong uniform static magnetic field for the polarization of nuclear spins.
In practice, there are several options:

- **permanent magnet**: low field applications (<0.3 T)
- **resistive magnet**: low field applications (<1.5 T)
- **superconducting magnet**: used for higher magnetic field strengths, e.g. the 9 T of today's high end research scanners.

The main magnet of the Magnetom Verio Scanner is only 6.5 tons!

Necessity of Gradients

To distinguish between the nuclei, the idea is to have a unique magnet field associated with each nuclei. This can be achieved by a continuous variation of the magnetic field dependent on the 3-D position of the nuclei.

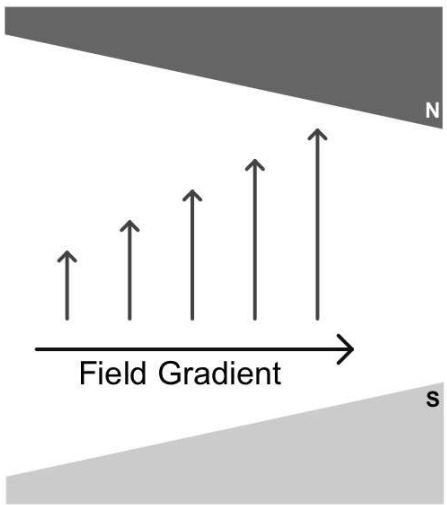


Figure: An example for a continuously changing magnetic field in 3-D

Note: The gradient strength in 3-D can be used for motion compensation. If the patient motion is known, the gradient field can be adjusted properly!

Components of an MR Scanner: Gradient System

The **magnetic field gradient system** is required to generate magnetic fields of well-defined and controlled spatial inhomogeneity as a function of the particular x, y and z coordinates in space. The gradient field is needed for signal localization in space.

- gradient strength: z.B. 45 mT/m (millitessla per meter) in the recent Tim Trio 3T
- rise time to ramp up gradient decides on quality of gradient (the smaller the rise time, the better); rise time today ˜1.0 ms

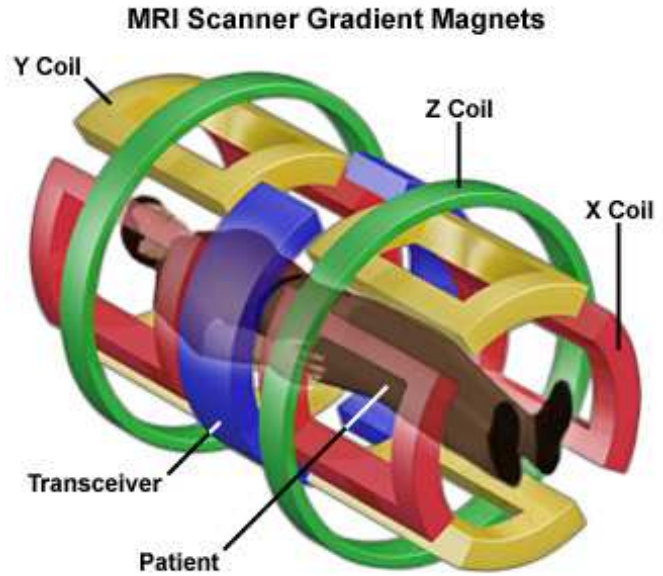


Figure: Principle Structure of the gradient system (image borrowed from here)

Components of an MR Scanner: RF System

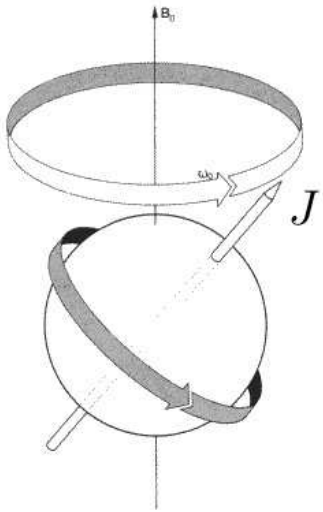
The **radio frequency system** has two components: transmission and receiver coil. In some systems transmission and receiver coils are identical. This is called a tranceiver coil.

- transmitter coil generates a rotating magnetic field for the excitation of a spin system
- receiver coil converts magnetic changes in electrical signals.

Figure: Sample of RF head and body matrix coil (image Siemens Medical Solutions)

Nuclear Magnetic Moments

- Pauli found out that particles have besides mass and charge another basic property: the **spin**
- the spin angular moment is denoted by $\boldsymbol{J}$
- with nuclear spin a microscopic magnetic field is associated because nuclei like protons carry electrical charges and rotate around own axes if spin is nonzero.
- the nuclear magnetic moment (or dipole moment) is denoted by $\mu \in \mathbb{R}^3$



- spin angular moment and magnetic moment are collinear, i.e.

$$\boldsymbol{\mu} = \gamma \cdot \boldsymbol{J}$$

where γ is the nucleus dependent gyromagnetic ratio.

Nuclear Magnetic Moments

- The theory of quantum mechanics shows that the magnitude of the magnetic moment is given by

$$||\boldsymbol{\mu}|| = \gamma \frac{h}{2\pi} \sqrt{I(I+1)}$$

  where $h = 6.626068 \cdot 10^{-34} \frac{m^2 kg}{s}$ is Planck's constant and I is the spin quantum number.
- I is an integer or half-integer: $I = 0, \pm\frac{1}{2}, \pm 1, \pm\frac{3}{2}, \ldots$
- To achieve macroscopic magnetism it is required to line up spin vectors and this can be enforced by a strong external magnetic field $\boldsymbol{B}_0$
- precession frequency ω_0 of $\boldsymbol{\mu}$ given by the Larmor equation $B_0 = ||\boldsymbol{B_0}||$ is

$$\omega_0 = \gamma B_0.$$

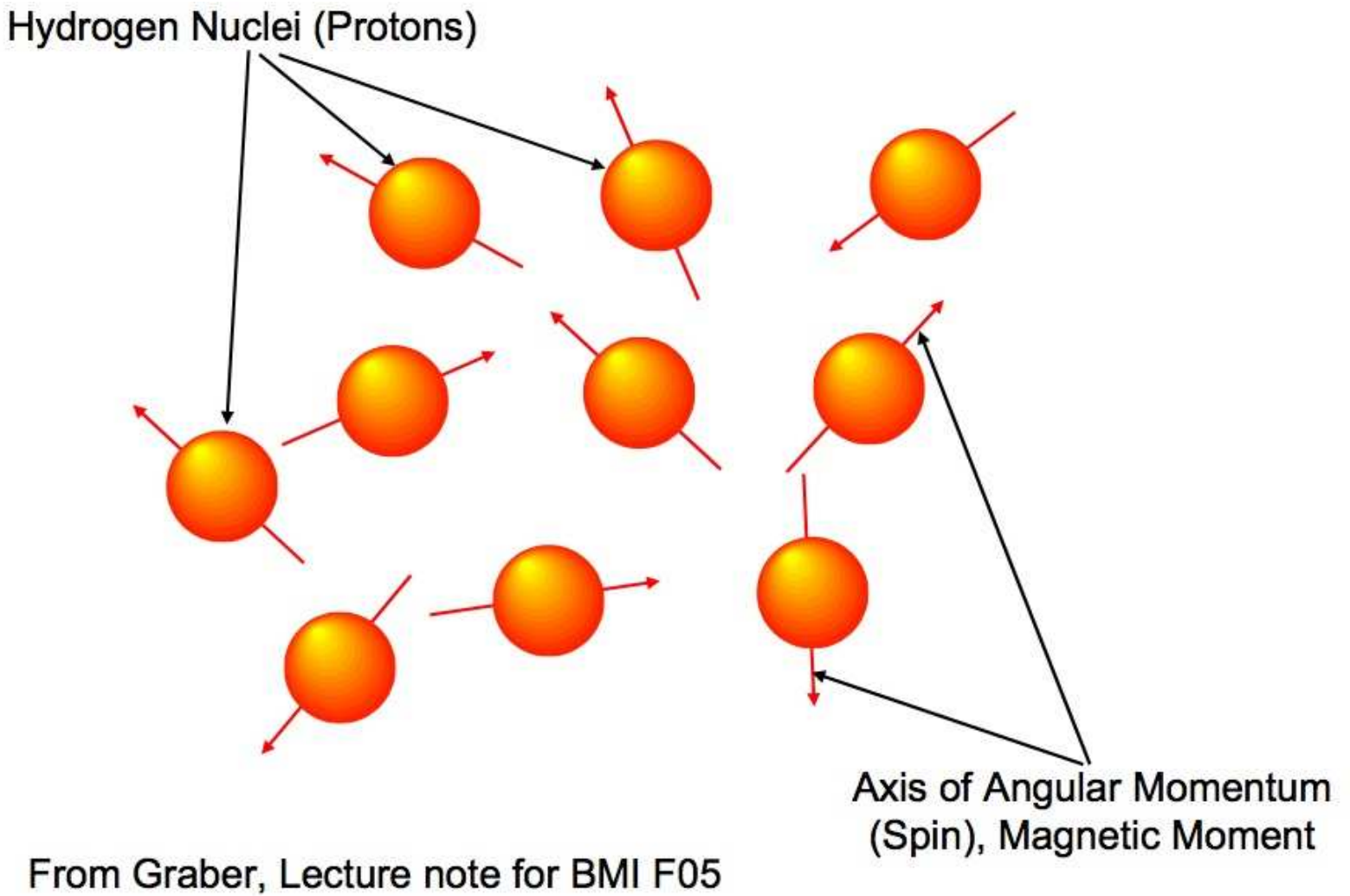
  The frequency ω_0 is called Larmor frequency.

Intensities in MR

Intensities in MR Images depend on resp. represent:

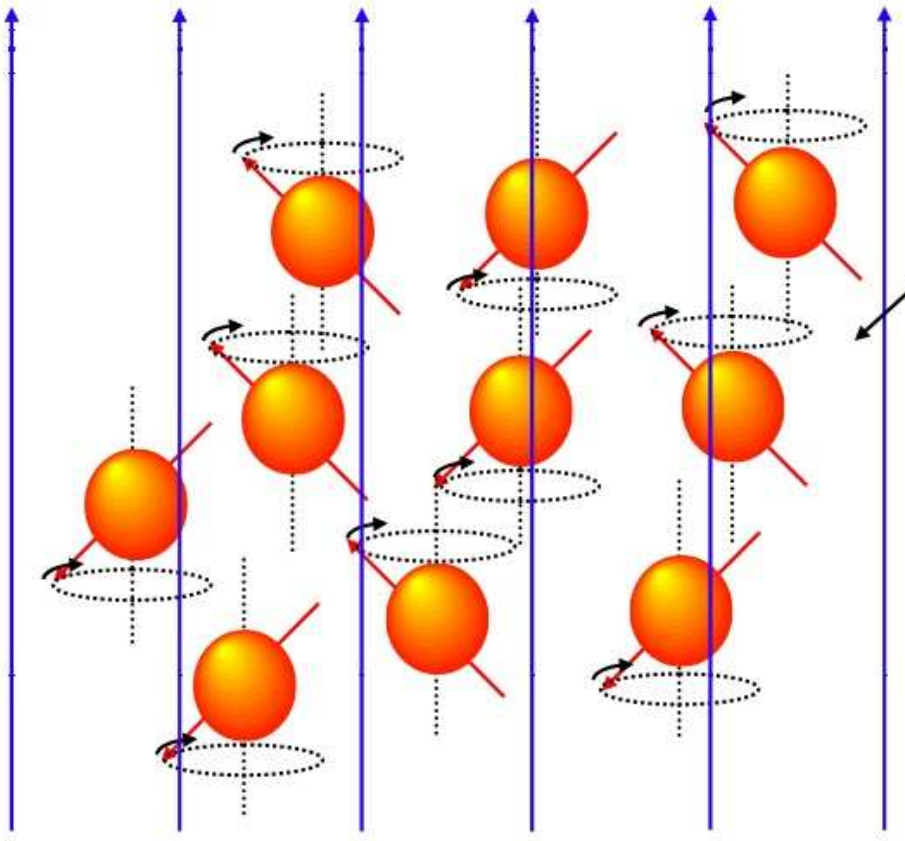
- spin density
- spin-lattice relaxation time (T1)
- spin-spin relaxation time (T2)
- molecular motion (diffusion, perfusion, flow)

MR Image Acquisition



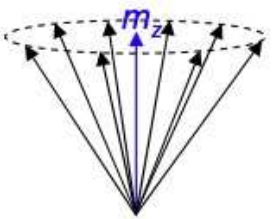
From Graber, Lecture note for BMI F05

MR Image Acquisition



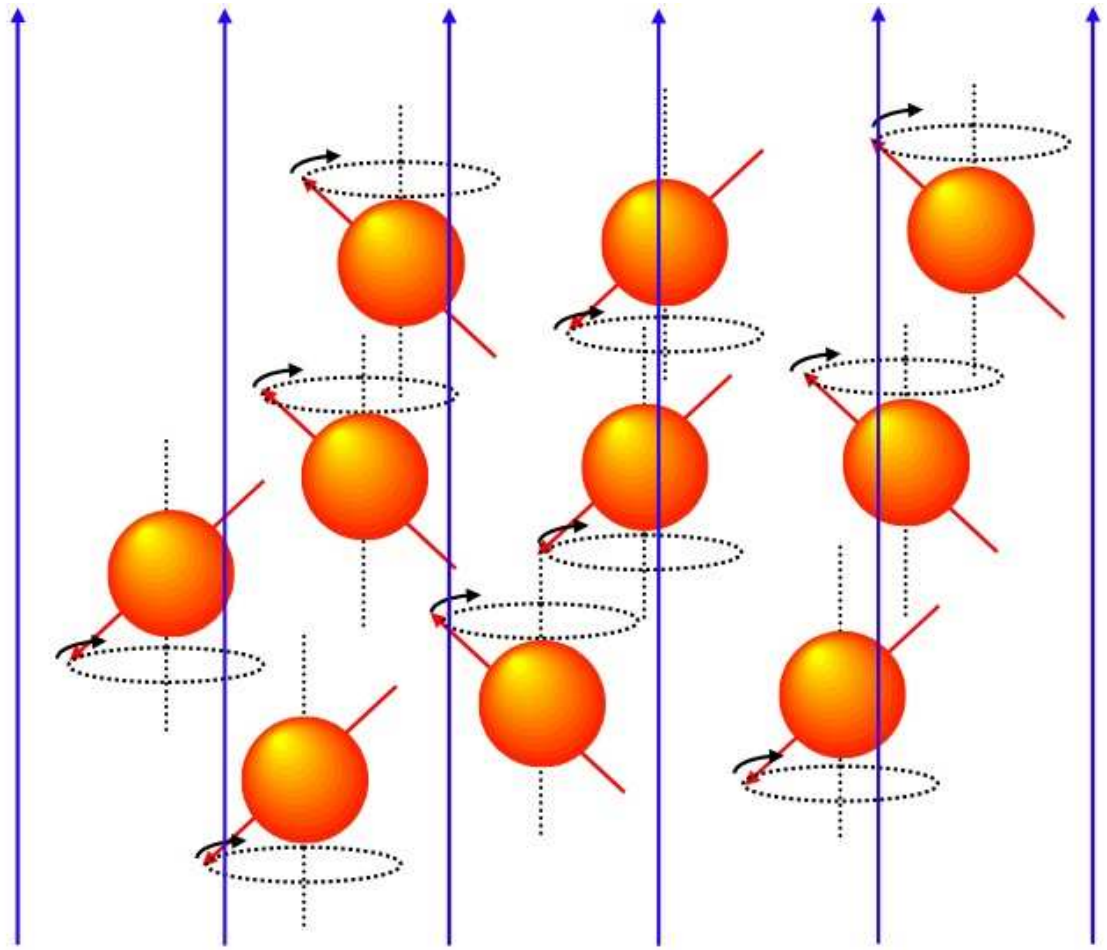
Spins **PRECESS** at a single frequency (w_0), but *incoherently !* they are not in phase, so that the sum of x-y components is 0, with net magnetization vector in z direction

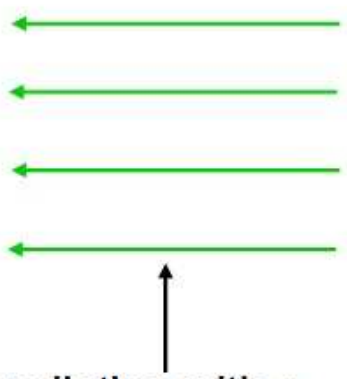
W0=\gamma B_0: Larmor freq.



From Graber, Lecture note for BMI F05

MR Image Acquisition





Irradiating with a rotating magnetic field B_1 of frequency w_0, causes spins to precess coherently, or in phase, generating a xy-component

Pros and Cons of MRI

Selected Pros

- patient care (MR uses radio frequencies $f = 10^6 - 10^8$ Hz, wavelength= 1-100 m, no X-ray!)
- high spatial resolution (50 μm and lower)
- excellent contrast resolution (discrimination of soft tissues)
- rich information about anatomical structure
- anatomical reference for functional modalities
- enables quantitative studies
- pre-, intra-operative guidance for intervention
- functional imaging modality (diffusion, perfusion, flow imaging)
- enables the usage of contrast agents
- magnetic field makes devices and especially hybrid scanners expensive.

Pros and Cons of MRI

Selected Cons

- expensive technology (investment and maintenance)
- inhomogeneities caused by radio frequency coil
- intensity inhomogeneities produce spatial changes in tissue statistics
- intensity-based segmentation fails due to inter- and intra-scan inhomogeneities
- inhomogeneities can change with different acquisition parameters from patient to patient and from slice to slice
- it is dangerous to approach the magnet with ferro magnetic objects, see for instance the images here and enjoy!

Diagnostic Medical Image Processing

Causes for Inhomogeneities

Definition

If slow and nonanatomic intensity variations are present in the image of one and the same tissue class, we talk about the presence of **intensity inhomogeneity** (IIH).

In other words, in an image with intensity inhomogeneities it can happen, for instance, that water molecules have different intensity values in the image domain. As a consequence no mapping of intensities to tissue classes is possible. Intensities allow for qualitative imaging.

Major reasons for intensity inhomogeneities in MR imaging are:

- non-uniform radio-frequency
- inhomogeneity of the static main field
- patient motion

Bias and Gain Fields

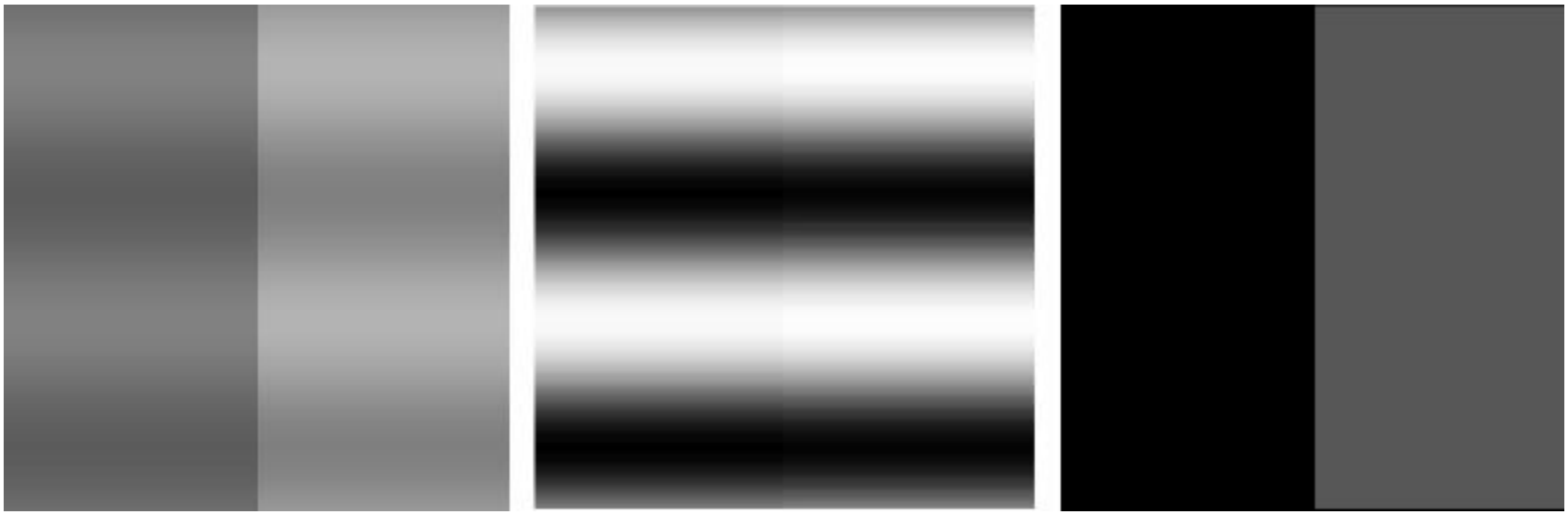


Figure: Observed image (left), gain field (middle) and ideal image (right) (image: W. Wells)

Inhomogeneities in MR

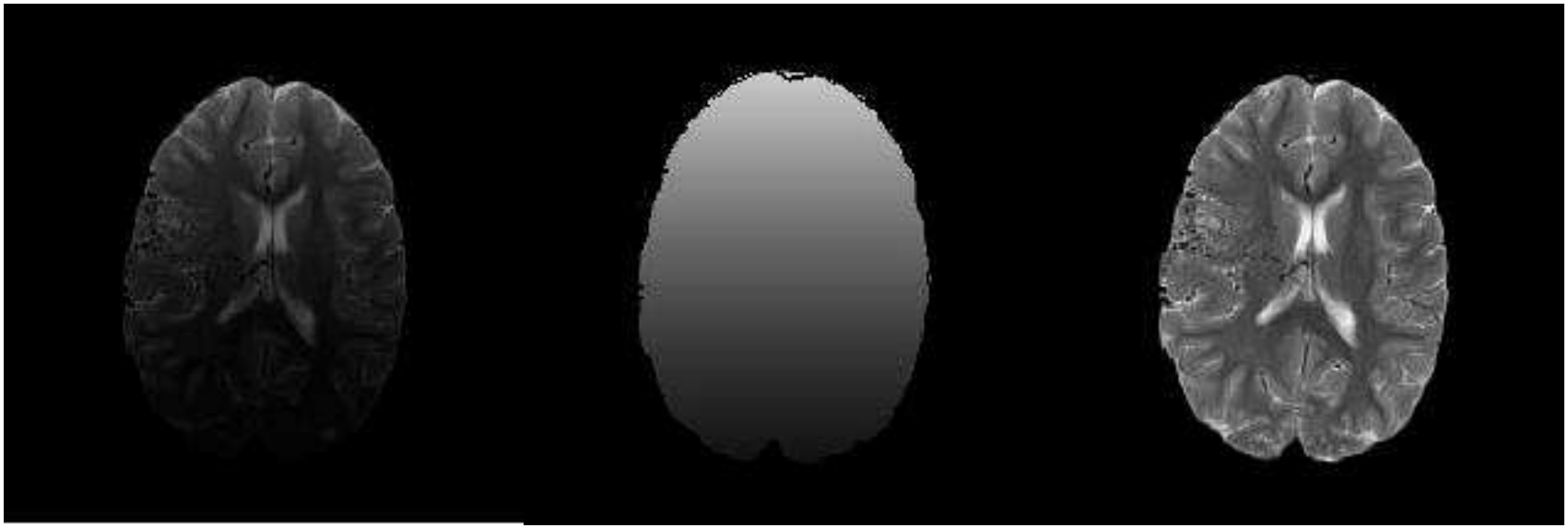


Figure: Original MR image (left), gain field (middle) and restored image (right)

Automatic Segmentation with/without Intensity Correction

In this figure the middle shows that naive segmentation using gradient information will produce inacceptable results. Only appropriate pre-processing implies a segmentation result as shown on the right.

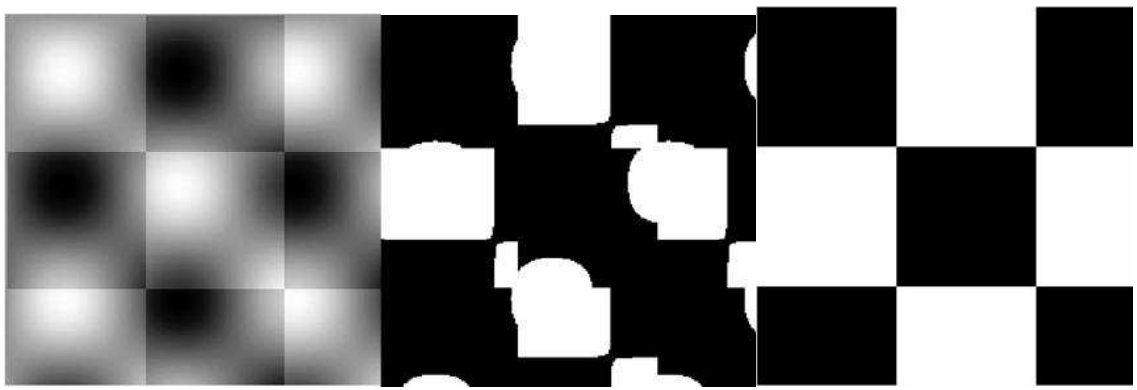


Figure: Biased checkerboard (left), segmentation result (middle) and ideal segmentation (right) (courtesy of W. Wells, Harvard Univ.)

Thresholding and Inhomogeneities

Even simple thresholding ideas fail due to intensity inhomogeneities.

Figure: Different MR images are binarized by the same threshold. It is obvious that heterogeneity maps identical tissue classes to different intensities (Florian Jäger, LME)

Diagnostic Medical Image Processing

Mathematical Models

Mathematical models for IIH are classified in three major categories:

1. **Low-frequency model:** It is assumed that IIH is caused by low-frequency components; the IIH map can be recovered by low-pass filtering.
2. **Hypersurface model:** IIH map is represented by a smooth (low-frequency) parametric function; the IIH map can be recovered by least-square-fitting (regression).
3. **Statistical model:** IIH map is represented by a stochastic process; the IIH map can be recovered dependent on the selected statistical model by parametric or non-parametric statistical estimation.

Mathematical Model

Definition

The **gain field** $b = [b_{i,j}]$ of is modeled as a multiplicative field which is applied to the ideal image $f = [f_{i,j}]$ pixelwise, i.e. all intensity values $f_{i,j}$ are multiplied by a spatially varying factor $b_{i,j}$:

$$g_{i,j} \quad = \quad f_{i,j} \cdot b_{i,j} + n_{i,j}, \quad \text{for} \quad i,j = 0,1,\ldots,N-1 \tag{1}$$

where $g_{i,j}$ denotes the observed intensity at grid point (i,j), and $n_{i,j}$ is additive Gaussian noise.

Note: The IIH correction is mostly applied to the product $b_{i,j}f_{i,j}$. Usually the Gaussian noise is eliminated by low-pass filtering, smooth model fitting, or regularization.

Mathematical Model

Definition

The **bias field** ${}^{\log}b = [{}^{\log}b_{i,j}] = [\log(b_{i,j})]$ results from the bias field by considering logarithms; in the absence of additive Gaussian noise this results in an additive model:

$$\log g_{i,j} \quad = \quad \log f_{i,j} + \log b_{i,j}, \quad \text{for} \quad i,j = 0,1,\ldots,N-1 \qquad (2)$$

Some researchers also incorporate additive noise in the above equation with logarithms, but this is not a proper noise model for practical applications.

Mathematical Model

- **Gain field correction:**
  If the gain field is known, the computation of the ideal image can be done pixelwise by:

$$f_{i,j} = \frac{g_{i,j}}{b_{i,j}} \tag{3}$$

- **Bias field correction:**
  If the bias field is known, the computation of the ideal image can be done pixelwise by:

$$\log f_{i,j} \quad = \quad \log g_{i,j} - \log b_{i,j}, \quad \text{for} \quad i,j = 0,1,\ldots,N-1 \tag{4}$$

Multiplicative vs. Log Additive Model

- early models for image inhomogeneity assumed a pure additive model, but
- additive effects are rarely observed in MRI
- smooth multiplicative inhomogeneity is in accordance to the physics of MRI imaging
- today multiplicative models are commonly accepted
- log additive model with Gaussian noise is not modeling the MRI acquisition specific noise properly

Diagnostic Medical Image Processing

Frequency Domain Filters

Design a high-pass filter that eliminates the low frequency bias field.

Frequency Domain Filters

Let us consider the high-pass filtering idea first by designing a filter in frequency domain.

- First the input observed image $g = [g_{i,j}]$ is Fourier transformed:

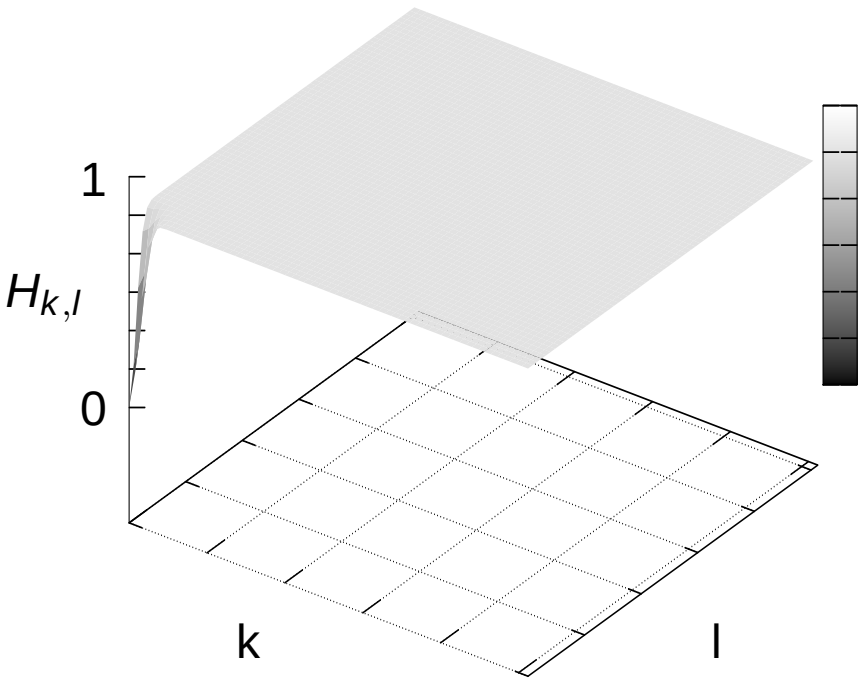
$$G = FT\{[g_{i,j}]\} \tag{5}$$

- A high-pass filter is defined in the discrete frequency domain by:

$$H_{k,l} = 1 - \beta \cdot e^{-\frac{k^2+l^2}{2\sigma^2}} \tag{6}$$

where β is a scaling factor that ensures that $H_{k,l} \geq 0$ for all $k, l = 0, 1, \ldots, N-1$ and σ^2 is closely related to the bandwidth of the filter-kernel.

Frequency Domain Filters



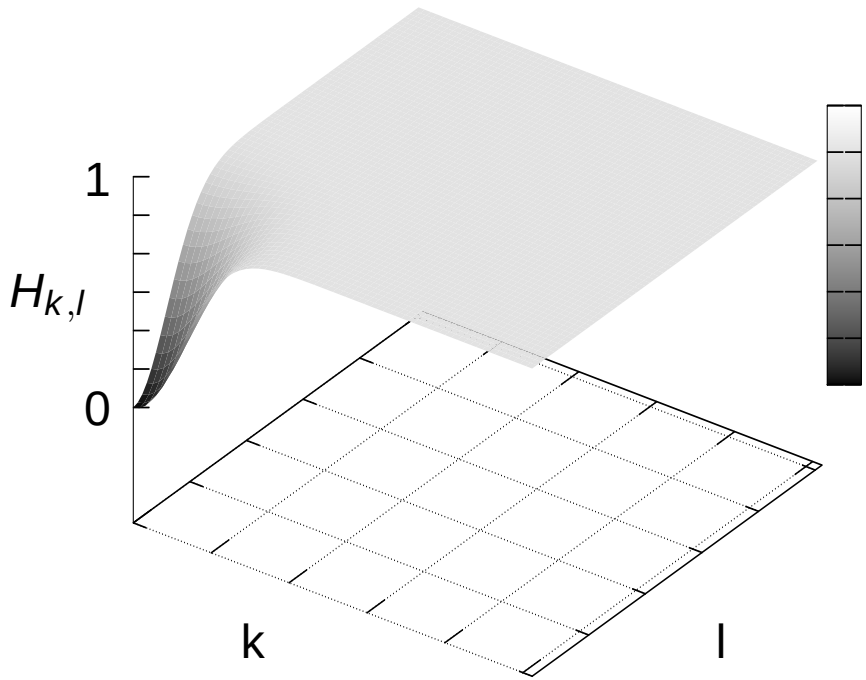


Figure: Qualitative plot of the shape of high-pass filter kernels (only positive frequencies shown due to symmetries) with $\sigma^2 = 60$ (left) and $\sigma^2 = 300$ (right)

Frequency Domain Filters

Using the convolution theorem, high-pass filtering is simply a multiplication in the frequency domain:

$$F_{k,l} = G_{k,l} \cdot H_{k,l} \tag{7}$$

The final output image f is obtained by computing the inverse Fourier transform:

$$f = FT^{-1}\{[F_{k,l}]\} \tag{8}$$

Diagnostic Medical Image Processing

Homomorphic Filtering

Filtering approaches assume IIH is

- an artifact with low frequencies, and
- the anatomic structures contribute to the high frequencies in the image

Elimination of image inhomogeneities can be done by low-pass filtering.

Homomorphic Filtering

Subtract low pass filtered image and normalize mean.

Homomorphic Filtering

Homomorphic filtering is applied to log-transformed images:

- low-pass filtering of the log-transformed image

$$[h_{i,j}] = \mathrm{LPF}([\log g_{i,j}])$$

  where LPF denotes a low pass filter, like mean or Gaussian filter

- IIH corrected log-transformed $\log f$ image results from the difference:

$$[\log f_{i,j}] = [\log g_{i,j}] - [h_{i,j}] + \mu$$

  where μ assures that IIH correction is mean preserving.

Exercise: Compute a formula to estimate μ.

Diagnostic Medical Image Processing

Homomorphic Unsharp Masking

Apply mean normalization.

Homomorphic Unsharp Masking

Homomorphic Unsharp Masking (HUM) is most of the simplest IIH correction methods and most commonly used approach.
HUM requires the computation of:

- global mean value μ of the intensity distorted image
- local mean values $\mu_{i,j}$ for each pixel considering a neighborhood.

If the multiplicative model is used, the estimated intensity corrected value $f_{i,j}$ is then computed pixel by pixel in the following manner:

$$f_{i,j} = g_{i,j} \cdot \frac{\mu}{\mu_{i,j}} \tag{9}$$

Homomorphic Unsharp Masking

A few remarks on homomorphic unsharp masking:

- This IIH-correction method relies on the basic assumption that the local mean in an image is equal to the global mean in the absence of IIH.
- Differences in local and global means are thus caused by the bias field only.
- This assumption only holds if the neighborhood used for computing the local mean contains a representative sampling of the tissue types in the image
- The size of the neighborhood has to be chosen carefully (experimental problem).

Diagnostic Medical Image Processing

Polynomial Surface Fitting

Approximate the low frequency bias field in the additive model by a multivariate regression polynomial and eliminate the bias by subtraction.

Polynomial Surface Fitting

The basic idea:

- logarithmic image is considered as a 2-D function, where the pixel coordinates (i, j) denote the sampling points and the intensities $\log(g_{i,j})$ the associated function values.
- fit a parametric smooth surface to the logarithm of intensity values
- estimate the parameters by minimizing the sum of squared differences of surface points and logarithms of image intensities.
- The resulting surface is then subtracted from the logarithmic image.

Polynomial Surface Fitting

According to our approach for spatial distortion correction, we define now a parametric mapping for intensity undistortion:

- we consider the image point at (i, j)
- the coefficients of the polynomials in i and j are denoted by $u_{k,l} \in \mathbb{R}$
- we assume separable base functions and thus require univariate base functions: $b_k : \mathbb{R} \rightarrow \mathbb{R}$, where $k = 0, \ldots, d$.

The polynomial that approximates the bias field is defined through the approximation:

$$g_{i,j} \approx \sum_{k=0}^{d} \sum_{l=0}^{d-k} u_{k,l} b_l(j) b_k(i) \tag{10}$$

(11)

Polynomial Surface Fitting

The resulting least square estimation problem is:

$$\hat{u}_{i,j} = \text{arg}min_{u_{k,l}} \sum_{i,j=0}^{N-1} ||g_{i,j} - \sum_{k=0}^{d} \sum_{l=0}^{d-k} u_{k,l} b_l(j) b_k(i)||^2 \quad (12)$$

This optimization problem can be solved by computing the SVD of the associated measurement matrix.
The final bias field estimate is:

$$b_{i,j} = \sum_{k=0}^{d} \sum_{l=0}^{d-k} \hat{u}_{k,l} b_l(j) b_k(i)$$

Exercise: Compute the measurement matrix for $b_l(j) = j^l$ and $b_k(i) = i^k$ and think about a proper scaling.

Diagnostic Medical Image Processing

Entropy and KL Divergence Minimization

Compute bias field that minimizes disorder in intensities.

Entropy and KL Divergence Minimization

The basic idea: Instead of fitting a parametric function with low frequencies by the minimization of the squared distance of the function to sampling values, i.e. intensities, we apply a statistical method:

- It is assumed that the probability density function of the intensities of the original unbiased image is multimodal.
- IIH causes intensity overlap and thus a “smearing” of the original probability density function. The peaks are no longer as sharp as they should be.
- Flattening brings the probability density function closer to a uniform density (increase of entropy).
- In terms of the statistical measures known from probability and information theory this means:
  - the bias field increases the entropy of the image
  - decreases the Kullback-Leibler divergence of the image probability density function of the image and the uniform density.

Entropy and KL Divergence Minimization

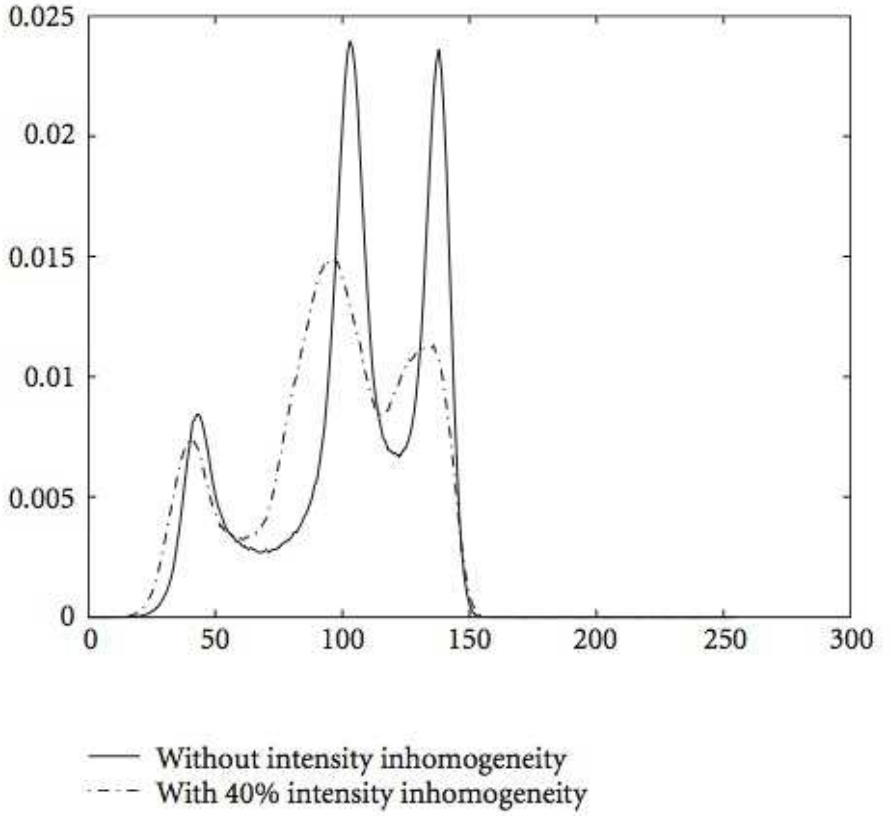


Figure: Histograms of brain tissue with IIH (dashed dotted line) and without bias (solid line)

Entropy and KL Divergence Minimization

Definition

The **entropy** of a discrete random variable X is defined by

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log p(x_i)$$

where the random measurements $x_1, x_2, \ldots, x_n$ underly the discrete probability density function $p(x_i) = p(X = x_i)$.

- entropy is maximum, if intensities are uniformly distributed
- entropy measures the amount of *disorder* in the image and it is mimimum for intensities with the least disorder.

Entropy and KL Divergence Minimization

Definition

The *Kullback–Leibler Divergence* (KL divergence) between two discrete probability density functions $\boldsymbol{p}$ and $\boldsymbol{q}$ is defined as:

$$KL(\boldsymbol{p}, \boldsymbol{q}) = \sum_{i=1}^{n} p(x_i) \log \frac{p(x_i)}{q(x_i)} \tag{13}$$

A few properties of the KL distance:

The KL divergence is a common similarity measure for probability density functions.

$$KL(\boldsymbol{p}, \boldsymbol{q}) \neq KL(\boldsymbol{q}, \boldsymbol{p}) \tag{14}$$

$$KL(\boldsymbol{p}, \boldsymbol{q}) \geq 0 \tag{15}$$

$$KL(\boldsymbol{p}, \boldsymbol{q}) = 0 \Leftrightarrow \boldsymbol{p} = \boldsymbol{q} \tag{16}$$

$$KL(\boldsymbol{p}, \boldsymbol{q}) \to 0 \quad \text{if} \quad p \to q \tag{17}$$

Entropy and KL Divergence Minimization

The algorithm for the estimation of the IIH

1. requires a parametric surface function that approximates the bias field, for instance, a bivariate polynomial and
2. the parameters are estimated such that
   - either the entropy of the resulting IIH corrected image is minimal or
   - the Kullback-Leibler divergence w.r.t to the uniform density is maximal.

Entropy and KL Divergence Minimization

Pros

- Can deal with rather complex and steep bias fields
- Uses only a few parameters, for the parametric representation of the bias field.

Cons

- Entropy minimization needs restrictions to avoid finding an all-white corrected image; this inhibits general applicability
- Difficult optimization task

Entropy and KL Divergence Minimization

Note:
In the Kullback-Leibler divergence based IIH correction, a reference probability density function besides the uniform density can be used to estimate the parameters of the bias field. In this case maximization is replaced by minimization. This allows the incorporation of prior knowledge to IIH correction.

Reference probability density functions can be generated from:

- reference data, for instance, images of a anatomic atlas data
- ideal data that are supposed to be unbiased, for instance, manually corrected intensities
- reference data that are acquired by costly high end MR scanners that provide superior image quality

Bias Field Estimation: Examples

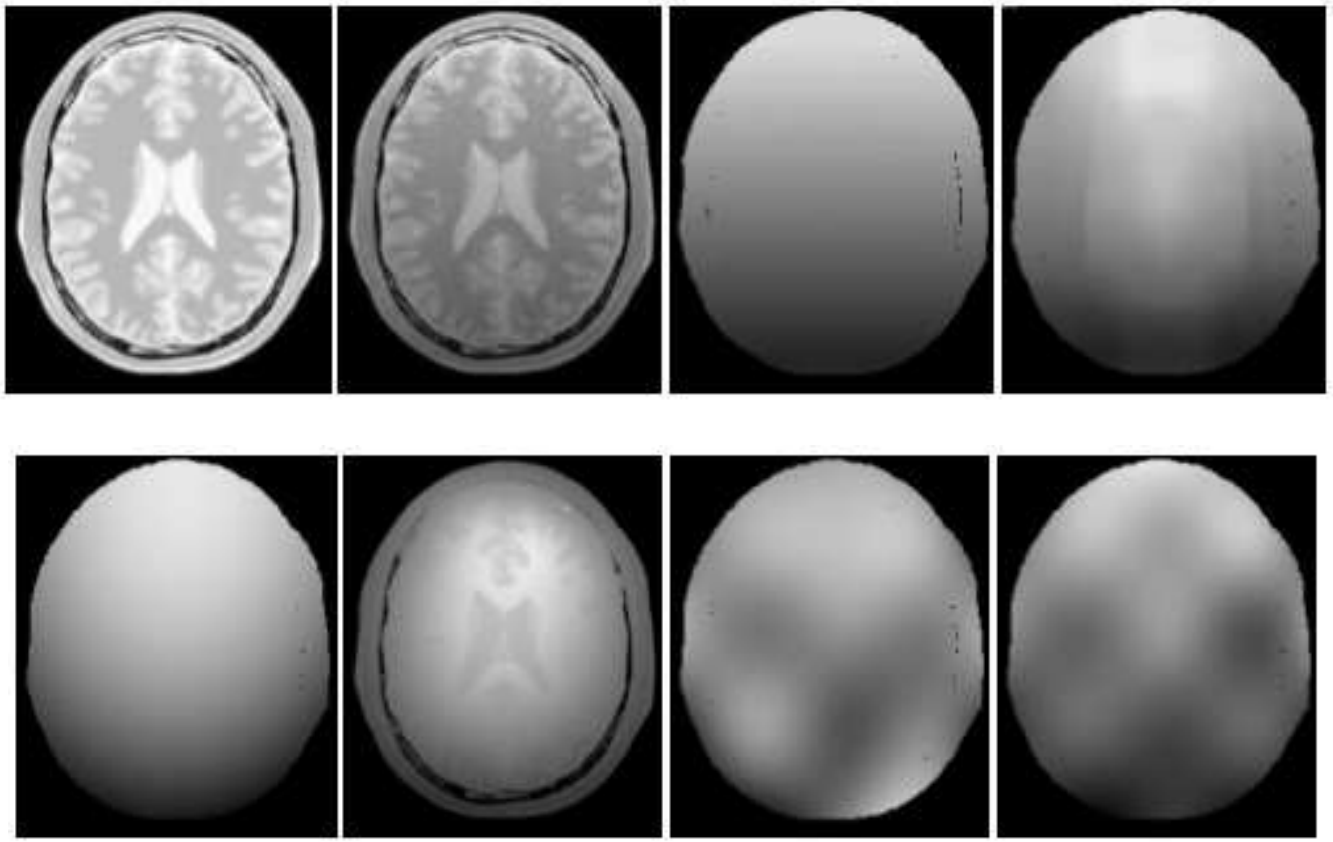


Figure: First row, from left to right: reference image, biased image (3% noise, 50% bias), original bias field, homomorphic unsharping mask. Second row: polynomial fit (degree 4), high pass filter, KL divergence with reference from original image, KL divergence with reference from high pass filter.

Diagnostic Medical Image Processing

Fuzzy C-means Clustering

Definition of the *fuzzy C-means objective function* (FCM) for partitioning the observations into N_c classes and allows one data point to belong to more than one classes:

$$J(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{N_c} \sum_{k=1}^{n} a_{i,k}^{d} ||x_k - c_i||^2 \tag{18}$$

- $c_1, c_2, \ldots, c_{N_c}$ are the prototypes of the clusters
- $x_1, x_2, \ldots, x_n$ are the data points, in our particular case the logarithms of ideal intensities

Fuzzy C-means Clustering

- $\boldsymbol{A} = [a_{i,k}]_{i=1,\ldots,N_c;k=1,\ldots,n}$ is called the **probabilistic partition matrix** satisfying the probability constraint:

$$\sum_{i=1}^{N_c} a_{i,k} = 1 \tag{19}$$

  for all samples $k = 1, 2, \ldots, n$.
- d is a weighting exponent.

Fuzzy C-means Clustering

Example

Let us assume

- three clusters: $i = 1, 2, 3$
- five data points: $k = 1, 2, 3, 4, 5$

For k-means clustering the partition matrix is:

$$[a_{i,k}] = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

For fuzzy C-means clustering the partition matrix is:

$$[a_{i,k}] = \begin{pmatrix} 0.5 & 0.7 & 0 & 0.1 & 0.6 \\ 0.3 & 0.2 & 1 & 0.9 & 0.3 \\ 0.2 & 0.1 & 0 & 0 & 0.1 \end{pmatrix}$$

Fuzzy C-means Clustering

There are a few drawbacks:

- the current objective function with the probabilistic assignment of data points to classes does not consider dependencies of neighboring data points (intuition: neighboring data points most probably belong to the same class)
- probabilistic approach required mutually independent intensities

The question now is, how can we incorporate dependencies of neighboring data points?

Fuzzy C-means Clustering

Idea: Extend FCM-objective function by regularization.

- regularization allows the
  - incorporation of *prior knowledge* and/or
  - introduction of *penalty terms*.
- add a term that biases the solution of the optimization problem towards a piecewise homogeneous labeling

Fuzzy C-means Clustering

A possible **regularized** objective function is:

$$J_R(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{N_c} \sum_{k=1}^{n} a_{i,k}^d ||x_k - c_i||^2 + \sum_{i=1}^{N_c} \sum_{k=1}^{n} \frac{\lambda}{\#\mathcal{N}_k} a_{i,k}^d \sum_{x_r \in \mathcal{N}_k} ||x_r - c_i||^2$$

where:

- $\mathcal{N}_k$ represents the particular set of neighbors of x_k.
- $\#\mathcal{N}_k$ is the cardinality of the considered neighborhood
- λ is a weighting factor.

Fuzzy C-means Clustering

Now we replace the logarithm of ideal intensity value x_k using $x_k = y_k - \beta_k$ and solve the optimization problem:

$$\{\hat{\boldsymbol{A}}, \hat{c}_i, \hat{\beta}_i\} = \text{argmin}_{\boldsymbol{A}, c_i, \beta_k} \sum_{i=1}^{N_c} \sum_{k=1}^{n} a_{i,k}^d ||x_k - c_i||^2 + \sum_{i=1}^{N_c} \sum_{k=1}^{n} \frac{\lambda}{\#\mathcal{N}_k} a_{i,k}^d \sum_{x_r \in \mathcal{N}_k} ||x_r - c_i||^2$$

subject to the probability constraint

$$\sum_{i=1}^{N_c} a_{i,k} = 1 \quad \text{for all} \quad k = 1, 2, \ldots, n$$

Fuzzy C-means Clustering

Membership Evaluation:

- The optimization regarding $a_{i,k}$ has to satisfy the aforementioned probability constraint, i.e. sum up to one for all k.
- For the incorporation of the probability constraint, we have to apply the Lagrange multiplier method

The extended objective function is:

$$J_R = \sum_{i=1}^{N_c} \sum_{k=1}^{n} \left(a_{i,k}^d D_{i,k} + \frac{\lambda}{\#\mathcal{N}_k} a_{i,k}^d E_{i,k} \right) + \eta \sum_{k=1}^{n} (1 - \sum_{j=1}^{N_c} a_{j,k}) \quad (20)$$

where

$$D_{i,k} = ||y_k - \beta_k - c_i||^2 \quad (21)$$

and

$$E_{i,k} = \sum_{(y_r - \beta_r) \in \mathcal{N}_k} ||y_r - \beta_r - c_i||^2 \quad (22)$$

Exercise: Compute the Lagrange multiplier η.

Estimator for Partition Matrix

The computation of the zero crossings of the gradient results in the following estimator for the partition matrix:

$$\hat{a}_{i,k} = \frac{1}{\sum_{j=1}^{N_c} \left(\frac{\#\mathcal{N}_k D_{i,k} + \lambda E_{i,k}}{\#\mathcal{N}_k D_{j,k} + \lambda E_{j,k}} \right)^{\frac{1}{d-1}}} \tag{23}$$

Cluster Prototype Update

$$\hat{c}_i = \frac{\sum_{k=1}^{n} a_{i,k}^{d} \left((y_k - \beta_k) + \frac{\lambda}{\#\mathcal{N}_k} \sum_{y_r \in \mathcal{N}_k} (y_r - \beta_r) \right)}{(1+\lambda) \sum_{k=1}^{n} a_{i,k}^{d}} \tag{24}$$

Bias Field Estimator

$$\hat{\beta}_k = y_k - \frac{\sum_{i=1}^{N_c} a_{i,k}^d c_i}{\sum_{i=1}^{N_c} a_{i,k}^d} \tag{25}$$

Fuzzy C-means Clustering: Image Examples

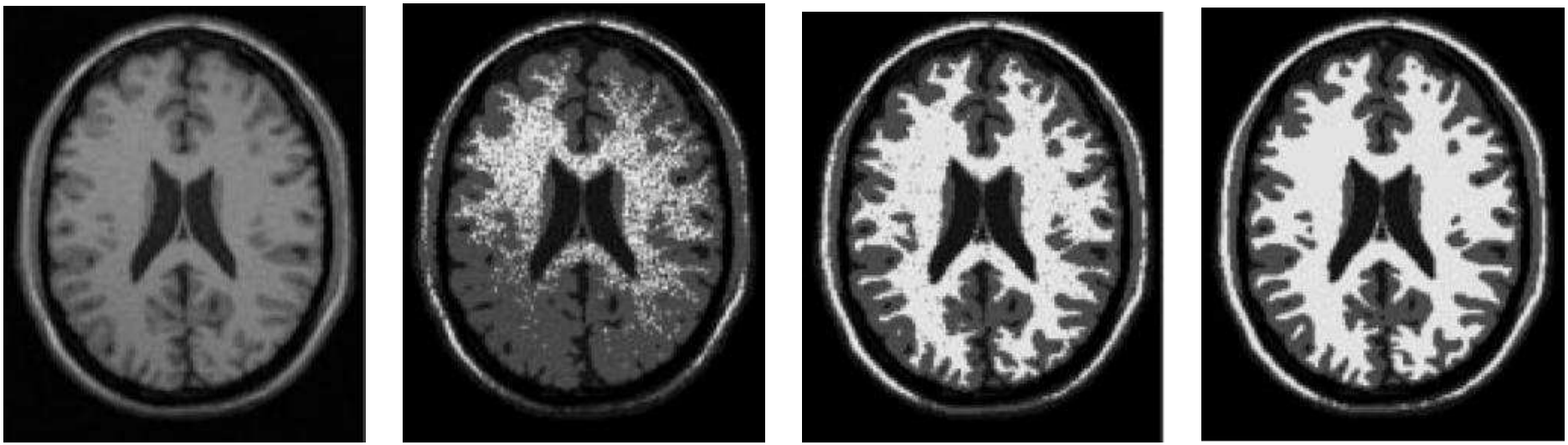


Figure: Comparison of segmentation results (from left to right): T1 weighted MR phantom image, fuzzy C-means algorithm without regularization, EM segmentation, fuzzy C-means with regularization [2]

[1] Example from paper by M.N. Ahmed, S.M. Yamany, N. Mohamed, A. A. Farag, T. Moriatry

[2] Example from paper by M.N. Ahmed, S.M. Yamany, N. Mohamed, A. A. Farag, T. Moriatry

Diagnostic Medical Image Processing

Probabilistic Model

Like in bias field correction using C-means clustering, we combine

- bias field correction with
- image segmentation

The **problem** we consider in the following is:
Compute the unbiased image where

- **tissue classes** of pixels, i.e. segmentation, **and**
- **bias field** are **unknown**.

Probabilistic Model

Core idea of the probabilistic approach:
Consider involved measurements as random measures that fit into the EM framework:

- observable random measurement: biased logarithmic intensity value
- hidden random measurement: tissue class for each pixel
- parameter estimation problem: computation of the bias field

This leads to an

incomplete data estimation problem

Probabilistic Model

The used probabilistic model consists of the following components:

- logarithmic intensities $x_{i,j} = \log g_{i,j}$ belonging to tissue class Γ are normally distributed

$$\begin{aligned} p(x_{i,j}|\Gamma;\beta_{i,j}) &= \mathcal{N}(x_{i,j};\mu_\Gamma+\beta_{i,j},\Sigma_\Gamma) \\ &= \frac{1}{\sqrt{|2\pi\Sigma_\Gamma|}}\exp\left(-\frac{1}{2}(x_{i,j}-\mu_\Gamma-\beta_{i,j})^T\Sigma_\Gamma^{-1}(x_{i,j}-\mu_\Gamma-\beta_{i,j})\right) \end{aligned}$$

  where

  - $x_{i,j}$: observed log intensity
  - Σ_Γ: covariance matrix of tissue class Γ
  - $\beta_{i,j}$: bias at point (i,j)
  - μ_Γ: mean log intensity of tissue class Γ

Mathematical Model

How can we prove that a certain random variable is Gaussian distributed?

- statistical testing: Kolmogorov-Smirnov test
- random variable is normally distributed by construction (for instance, principal components)
- assumption is approved by experiment (it simply works!)

Mathematical Model

Furthermore the probabilistic model consists of:

- prior probability of tissue class, i.e. without considering any observation:

$$p(\Gamma), \quad \text{for} \quad \Gamma = 1, 2, \ldots, N \tag{26}$$

- prior density of bias field:

$$p(\beta_{i,j}) \quad = \quad \mathcal{N}(\beta_{i,j}; 0, \Sigma_\beta) \tag{27}$$

  i.e. the pdf for the bias field is independent from (i, j).
- elimination of unknown tissue class Γ by marginalization:

$$p(x_{i,j}; \beta_{i,j}) \quad = \quad \sum_{\Gamma=1}^{N} p(\Gamma) p(x_{i,j}|\Gamma; \beta_{i,j}) \tag{28}$$

  i.e. a convex combination of density function.

Mathematical Model

Due to the fact that $p(\beta_{i,j})$ is assumed to be Gaussian, the bias field is now considered as a random variable. We do estimate the bias field $\boldsymbol{\beta} = [\beta_{i,j}]$ by the maximization of posteriors given the logarithmic image $\boldsymbol{x} = [x_{i,j}]$.

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= \operatorname{argmax}_{\boldsymbol{\beta}} p(\boldsymbol{\beta}|\boldsymbol{x}) \\
&= \operatorname{argmax}_{\boldsymbol{\beta}} \frac{p(\boldsymbol{\beta})\, p(\boldsymbol{x}|\boldsymbol{\beta})}{p(\boldsymbol{x})} \\
&= \operatorname{argmax}_{\boldsymbol{\beta}} p(\boldsymbol{\beta})\, p(\boldsymbol{x}|\boldsymbol{\beta}) \\
&= \operatorname{argmax}_{\boldsymbol{\beta}} \left(\log p(\boldsymbol{\beta}) + \log p(\boldsymbol{x}|\boldsymbol{\beta})\right) \qquad (29)
\end{aligned}$$

Both logarithmic densities result from the probabilistic modeling we have introduced so far.

Mathematical Model

Commonly applied (but not necessarily right) simplification:
Intensities are assumed to be mutually independent, i.e.

$$p(\boldsymbol{x}|\boldsymbol{\beta}) = \prod_{i,j} p(x_{i,j}|\beta_{i,j}) = \prod_{i,j} \sum_{\Gamma=1}^{N} p(\Gamma) p(x_{i,j}|\Gamma, \beta_{i,j}). \tag{30}$$

The estimation of the bias field can be done iteratively, but the current model includes hidden variables. In total, we have to estimate the following parameters using the EM:

- bias field $\boldsymbol{\beta}$
- covariance Σ_β
- priors $p(\Gamma)$
- means $\boldsymbol{\mu}_\Gamma$ and covariances Σ_Γ.

Mathematical Model

Once all the parameters are estimated, the segmentation result is required. The final tissue class of each pixel can be estimated by the maximization of the posteriors:

$$\begin{aligned} \hat{\Gamma} &= \operatorname{argmax}_{\Gamma} p(\Gamma | x_{i,j}; \beta_{i,j}) \\ &= \operatorname{argmax}_{\Gamma} (\log p(\Gamma) + \log p(x_{i,j} | \Gamma; \beta_{i,j})) \end{aligned}$$

Remark on Bias Field Modeling

In practice the bias field is not estimated for all components $\beta_{i,j}$, but usually approximated by a parametric function:

$$\beta_{i,j} = \sum_{k=0}^{M} \theta_k \ \phi_k(i,j)$$

where we have $\theta_k \in \mathbb{R}$ and ϕ_k are proper base functions.
The bias field estimation thus reduces to the computation of the θ_k's.

Exercise: Compute the iteration schemes to estimate the bias field by the usage of EM and enjoy the math!

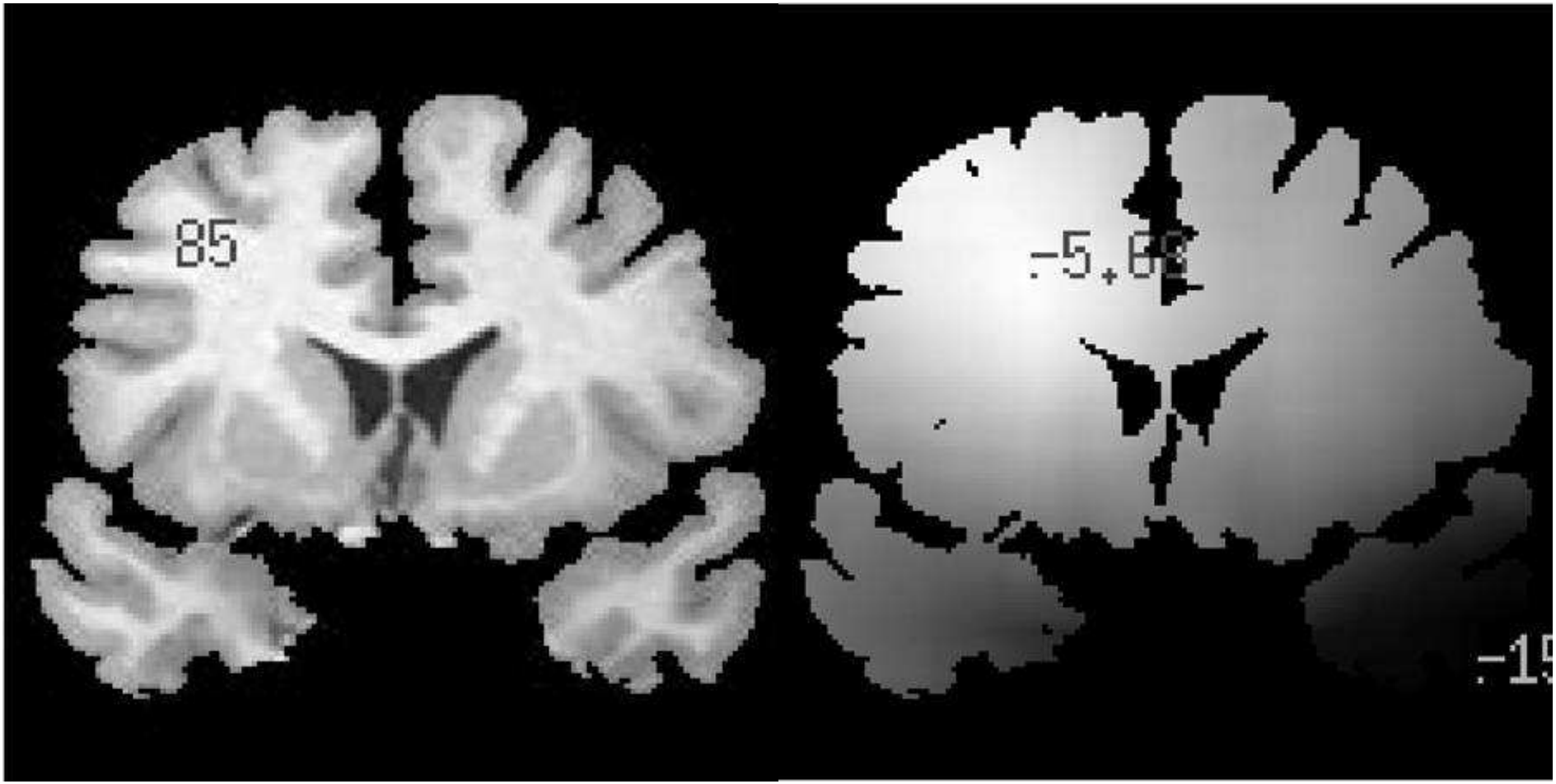


Figure: Original MRI (left) and the estimated bias field (right) using the EM technique (Courtesy of W. Wells)

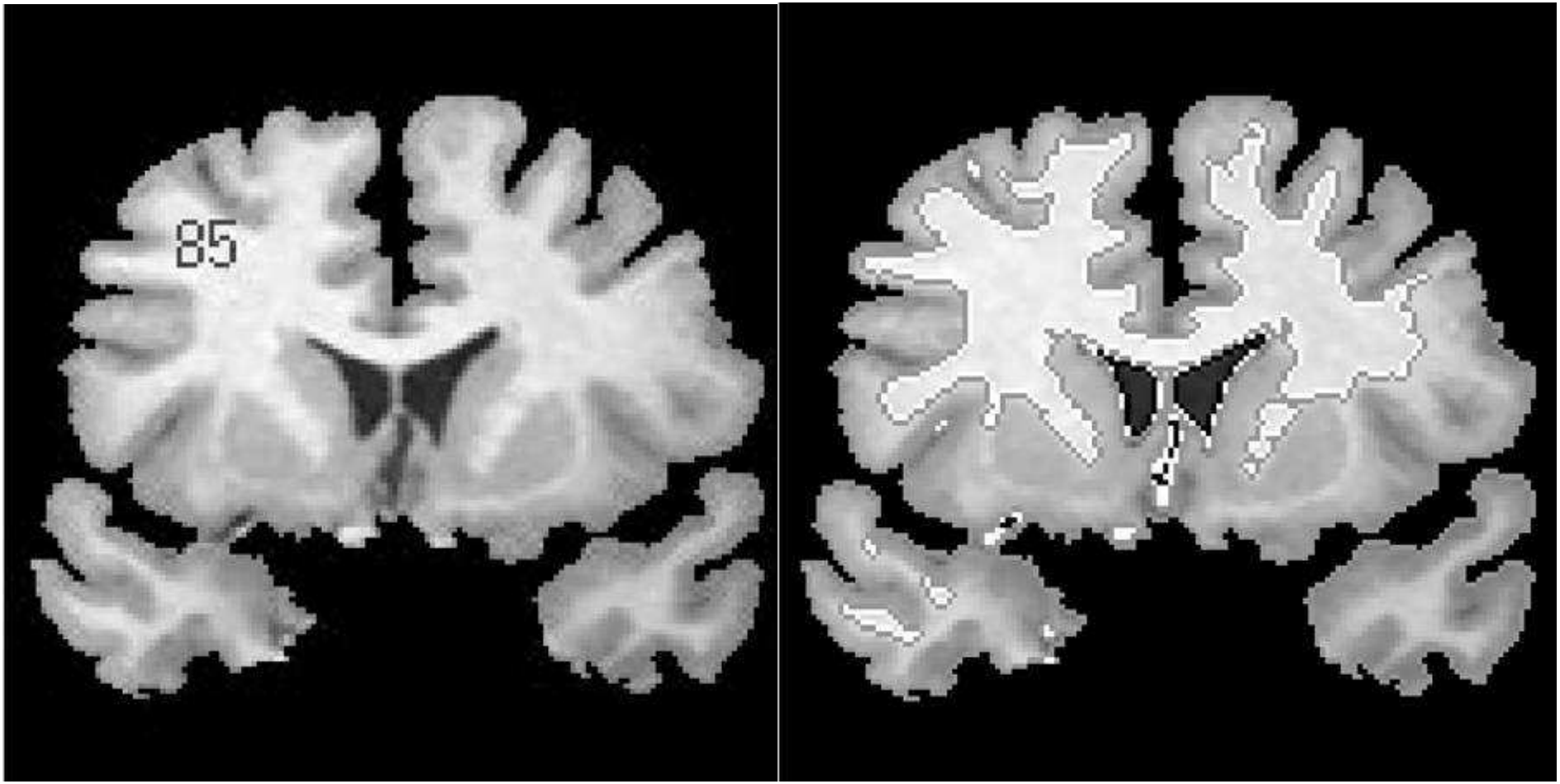


Figure: Original MRI (left) and the result of segmentation (right) (Courtesy of W. Wells)

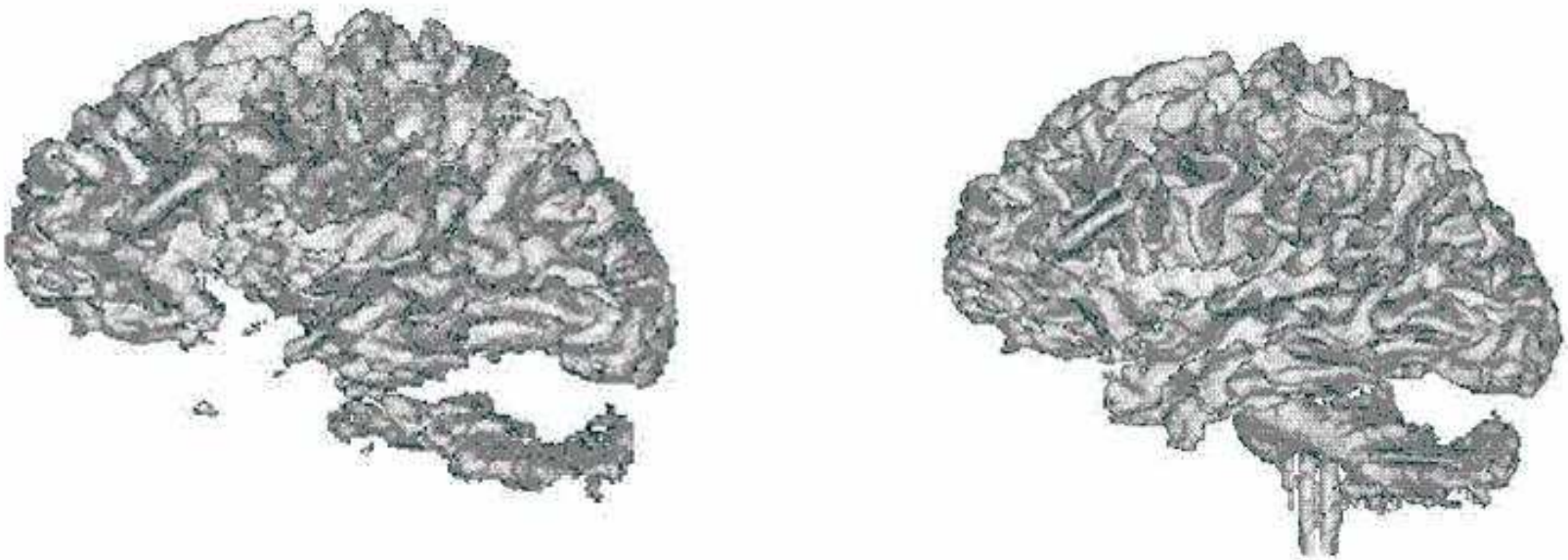


Figure: White matter surface computed by conventional and adaptive segmentation (Courtesy of W. Wells)

Diagnostic Medical Image Processing

Further Applications of IIH

- X-ray imaging: correction of the Heel Effect
- endoscopy/retina imaging: correction of heterogeneous illumination
- ultrasound imaging: correction of signal decay with distance from probe and of shadows

Further Applications of IIH

Bias correction in retina image processing:

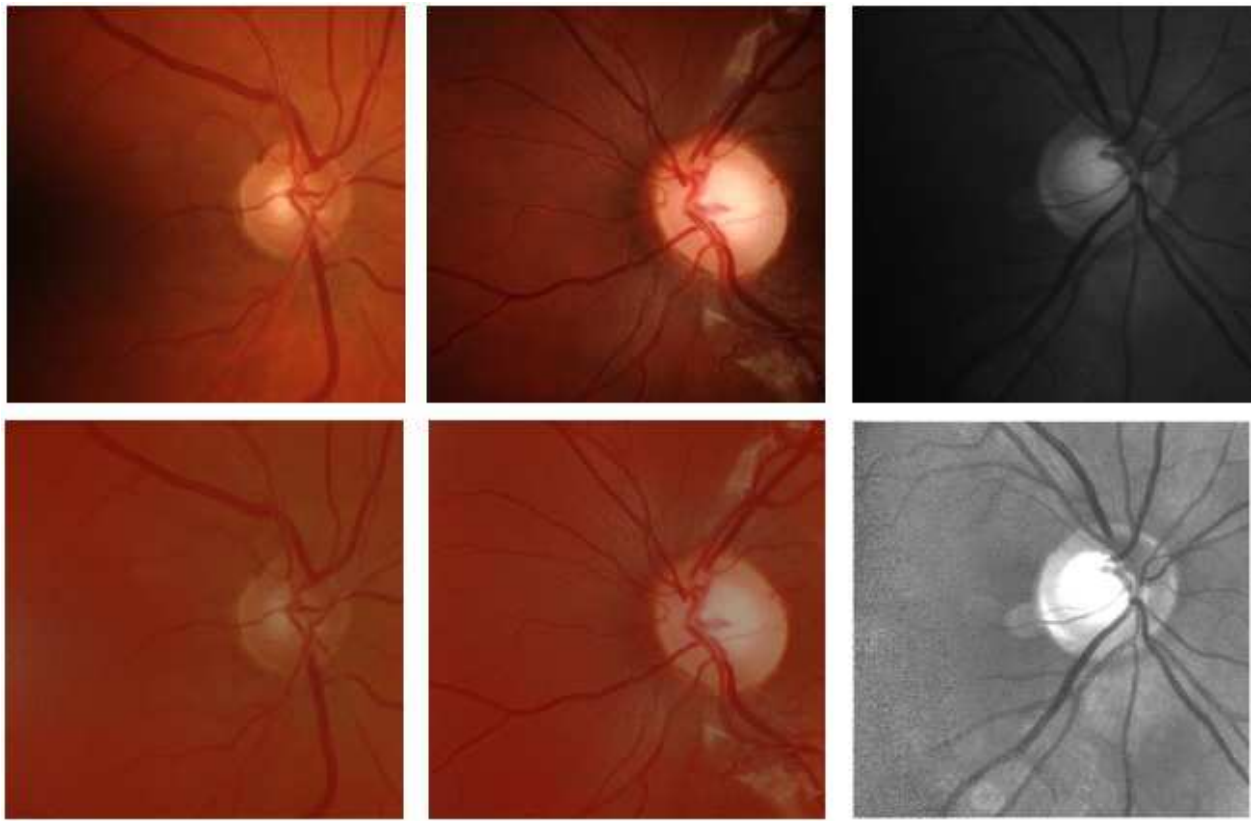


Figure: Bias correction examples: retina images with heterogeneous illumination (upper row), bias corrected images (surface fitting method, degree 4 polynomials, lower row).

Further Applications of IIH

Bias correction in ultrasound imaging:

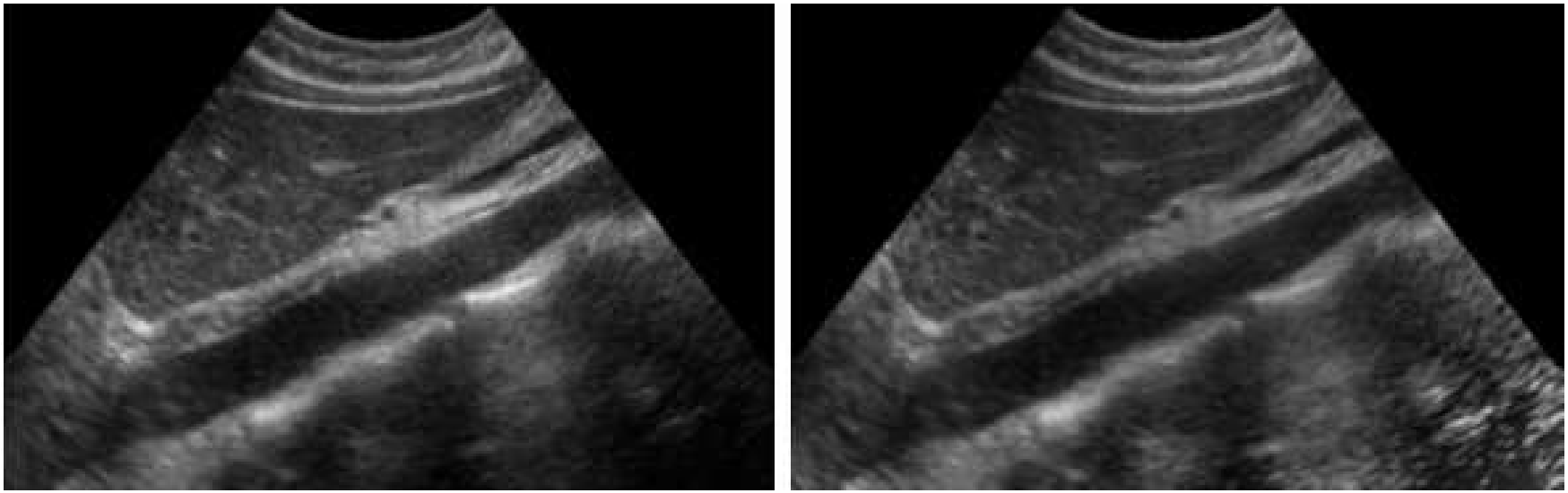


Figure: Bias correction examples: ultrasound image with decreasing signal from top to bottom (left), bias corrected image (surface fitting method, degree 1 polynomial, right)

Diagnostic Medical Image Processing

Take Home Messages

- gain and bias correction are crucial not only but especially in MRI
- we have discussed the following classes of retrospective IIH correction:
  - filtering
  - surface fitting
  - histogram based
  - segmentation
- incomplete data estimation

Diagnostic Medical Image Processing

Further Readings

- A worth reading survey paper on the most recent state-of-the-art algorithms for intensity correction is:
  Zujun Hou: A Review on MR Image Intensity Inhomogeneity Correction International Journal on Biomedical Imaging, Vol. 2006, Article ID 49515, p. 1-11
- Behiels, G.; Maes, F.; Vandermeulen, D.; Suetens, P.: Retrospective heel effect correction in conventional radiography IEEE Workshop in Mathematical Methods in Biomedical Image Analysis, 2001, Pages: 87 - 94

Further Readings

- Recommended book on EM algorithm:
  G.J. McLaughlan, T. Krishnan, The EM Algorithm and Extensions, Wiley Series in Probability and Statistics, John Wiley & Sons, 1998
- Segmentation of MR Images:
  - Homepage of Sandy Wells
  - paper on adaptive segmentation of MR images (EM based)
  - paper on adaptive segmentation of MR images (fuzzy C–means)
- Details on the curse of dimensionality can be found in:
  The Elements of Statistical Learning