Michael Treml, BSc

Bayesian Analysis of Magnetic Resonance Image Reconstruction

MASTER THESIS

For obtaining the academic degree
Diplom-Ingenieur

Master Programme of
Technical Physics

Supervisor:
Univ.-Prof. Dipl.-Phys. Dr.rer.nat. Wolfgang Von der Linden
Institute of Theoretical and Computational Physics

Graz, August 2014

Abstract

The aim of this thesis is to study the behaviour of regularized image reconstruction of parallel MR data subjected to Gaussian noise for different grades of undersampling. As reconstruction method sensitivity encoded parallel imaging (SENSE) with a Tikhonov image prior is used. This reconstruction approach is one of the simplest, but has the advantage that many results can be obtained analytically. For this type of image reconstruction the question of how to find the optimal regularization parameter is answered from the Bayesian point of view and an equation to extract the optimal parameter is presented. Two methods - the pseudo replica method and MCMC sampling - are employed to generate samples in image space from a single MR-dataset in order to gain information about the noise statistics in image space. For both sampling methods the analytic expressions for the posterior PDF are computed and the general Rician distribution for magnitude images of correlated complex Gaussian random variables is derived. Furthermore, a relative statistical image quality measures which use the covariance matrix of the image samples is applied to datasets with different degrees of undersampling. It is shown that pixel correlation effects can not be neglected for undersampled data without justification.

Keywords:
Bayesian probability theory, signal power prior, magnetic resonance imaging, parallel imaging, SENSE, Tikhonov regularization, MAP solution, parameter estimation, MCMC, pseudo replica method, Rician distribution

Special acknowledgements

This thesis was supported by Co-Supervisor Univ.-Prof. Dipl.-Ing. Dr.techn. Rudolf Stollberger from Institute of Medical Engineering of Graz University of Technology in context of the project "Magnetic Resonance Imaging" at SFB Research Center for Mathematical Optimization and Applications in Biomedical Sciences funded by a research grant of the FWF - Austrian Science Fund .

Danksagung

Diese Diplomarbeit wurde im Jahr 2014 am Institut für Theoretische Physik und Computational Physics sowie am Institut für Medizintechnik an der Technischen Universität Graz durchgeführt. Besonderer Dank gebührt meinem Betreuer Prof. Wolfgang Von der Linden und meinem Co-Betreuer Prof. Rudolf Stollberger sowie Dipl.-Phys. Matthias Schlögl, der mich mit großer Geduld in das Gebiet der MR-Bildrekonstruktionsverfahren eingeführt hat. Darüberhinaus danke ich meiner Familie - insbesondere meinen Eltern Maria und Josef - für die Unterstützung, Geduld und Motivation, die sie mir im Laufe meines Studiums entgegengebracht haben. Meinen StudienkollegInnen danke ich für die gute Zusammenarbeit und die bereichernden Erfahrungen, die sie mir neben dem Studium geschenkt haben. Mögen die im Laufe meines Studiums erlangten Fähigkeiten zum Wohle aller dienen.

Graz, September 8, 2014

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Das in TUGRAZonline hochgeladene Textdokument ist mit der vorliegenden Masterarbeit identisch.

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master's thesis.

Datum/*Date*	Unterschrift/*Signature*

Contents

List of Figures

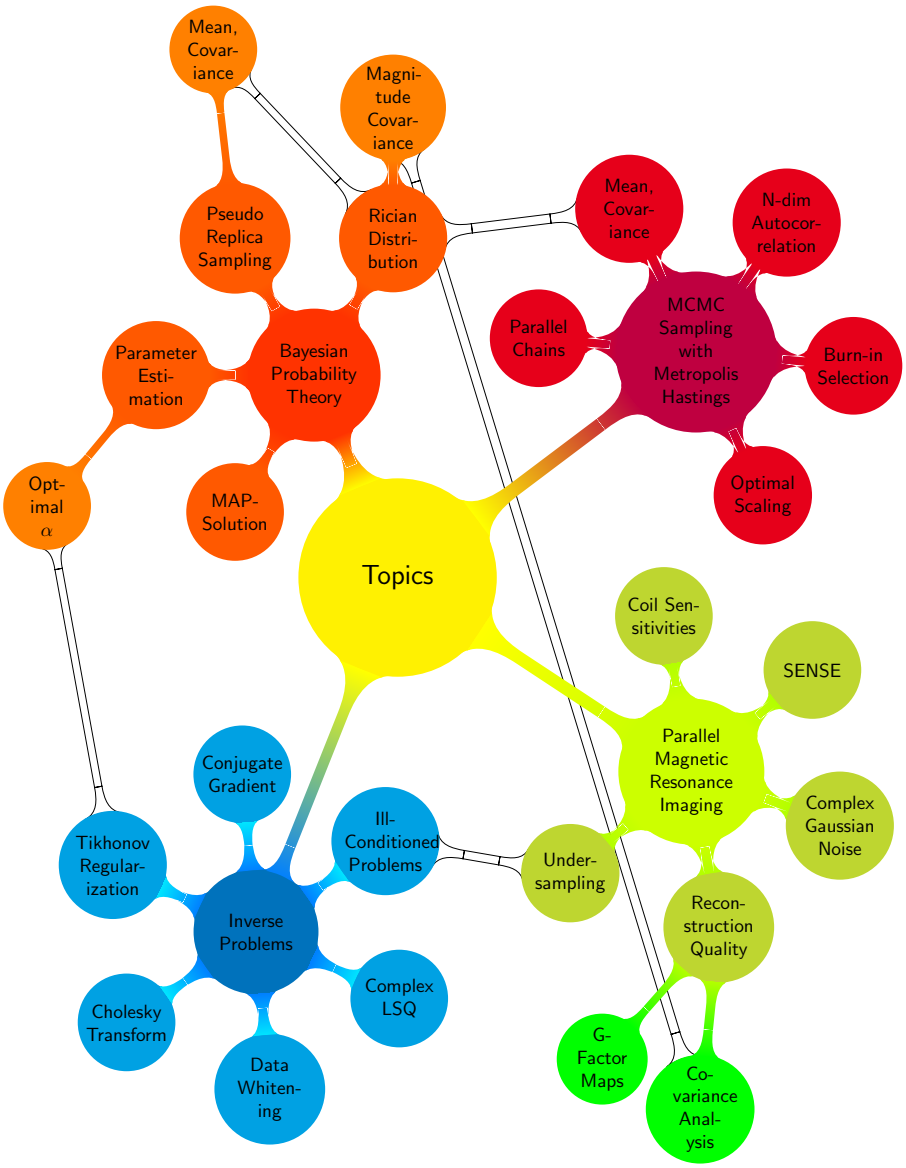


Figure 1: Mindmap of Contents

1 Introduction

MR scans are a common examination technique in today's clinical practice. MRI technology considerably simplifies medical diagnosis and helps to detect health problems in various cases. Unfortunately - due to motion artifacts - the examination of certain body parts like the gastrointestinal system, the lung or even the heart is either difficult or requires long scanning times. Another fact is that MR-scanners and their maintenance are very costly. From the health economical point of view it is therefore desirable to give as many patients as possible access to this medical service. Hence reducing the time per scan can increase comfort for patients as well as it can be beneficial for the health system. However, it raises a main question:

(1) *"How far can scanning time be reduced such that quality and reliability of the images can still be guaranteed?"*

To deal with this problem several parallel imaging techniques have been developed such as Sensitivity Encoding For Fast MRI (SENSE). To obtain reliable images a certain minimum resolution is necessary. In order to achieve the desired resolution the signal from a certain number of k-space points needs to be acquired. Due to the excitation and read-out sequence design one k-space data point takes a certain minimum time to be measured. That requires a certain minimum total scanning time which can be up to an hour or more. Furthermore also the noise in k-space plays an important role. Even though the resolution might be sufficient, very noisy data due to low signal-to-noise-ratio (SNR) might be amplified in certain regions during image reconstruction. This can result in totally useless images with bad (or misleading) diagnostic information. Hence a high SNR of k-space data is also necessary. Usually this is achieved by scanning the object paralleley with multiple receiver coils such that the effects of noise are reduced due to averaging. As such the described setup provides reliable images but will not give any speedup because in each coil still the whole k-space is sampled.

That is where another idea comes into play: Only half of the k-space is sampled which speeds up the scan by the factor 2. If the undersampled data received at each coil was going to be reconstructed individually, backfolding-artifacts would appear in the image because the sampling theorem gets violated. That is due to the properties of the Fourier transform. A certain spatial resolution imposes a minimal sampling stepsize in k-space. Given the fact that the spatial sensitivities of the receiver coils are known, the trick now is to reconstruct the image collectively by combining the data of all coils. The knowledge of the spatial dependencies of the coils can be used to remove the backfolding artifacts from the reconstructed image. If the parallel coils were perfectly linear independent (and hence contain no redundancy) theoretically the scanning time can maximally be reduced by the factor equal to the number of parallel coils without leading to any backfoldings artifacts. Techniques that perform MRI image reconstructions in that way are called parallel MRI and they are widely used by radiologists in today's clinical practice.

Numerous more advanced techniques have been developed to reduce scanning time beyond the bounds of the sampling theorem, e.g. compressed sensing in cardiovascular MRI. They are mostly based on the common idea to use prior information about some structure of an MR image like smoothness. The loss of information in data acquisition for faster scanning and the decrease in SNR is balanced by regularization terms within image reconstruction. This approach has been widely accepted in the scientific community but has not made its step into clinical application yet. The reason for that is that the quality of the regularized

and reconstructed image (or a certain region of interest (ROI)) is not known. A structure in the ROI subject to diagnosis might vanish due to excessive regularization (overfitting) and a possible pathology could become invisible. To avoid situations like that we take a look at the character of the image reconstruction which is performed solving an optimization problem with data-fidelity and regularization term. The latter is weighted with a certain constant which is usually chosen empirically to suite expectations best. The question that one is confronted with here is:

(2) *"What is the optimal (non-empirical) regularization parameter based only on the information we have about the problem?"*

To answer this question it would be convenient to have a profound derivation of how to find the optimal regularization parameter, only with respect to given information (model of measurement process, noise statistics and type of regularization). An equation can be derived by applying Bayesian probability theory that makes it possible to compute the optimal parameter from raw MRI data before performing a Tikhonov regularized SENSE reconstruction. This will help to put the choice of the regularization parameter on a profound statistical basis. Still reconstructed images can have highly biased ROIs due to noise amplification. Also for undersampling beyond the Nyquist boundary using regularizations noise amplification artifacts might still appear. The question a radiologist is going to face in such a case is:

(3) *"How reliable is a structure in the ROI?"*

Therefore it is desirable to have a mathematical precisely defined quality measure that assesses the quality of the ROI of a single scan. Since a measure of that kind is relative to the chosen regularization and the forward model with its noise characteristics, there can never exist an absolute measure. That becomes clear by looking at the following example: If the regularization favors leopard-patterns the quality of the reconstruction can be 100% but the image is still nonsense when looking at human MRI data. However, within the given assumptions (noise statistics, MRI model, type of regularization) it would be still very helpful for a radiologist to know the relative quality of the reconstruction in order to decide whether he is diagnosing the patient or rather sending him back to do another fully sampled scan. Such a measure is already found in literature [20] as (geometry) g-factor maps and it can be computed pixelwise from a single dataset by using the pseudo replica method. It has one fundamentally drawback: it lacks the consideration of pixel correlation effects which are of great importance in practice. In MR images that have been reconstructed from undersampled data the pixels in a ROI are not fully uncorrelated anymore which can lead to fatal errors in quality assessment by using diagonal elements of the covariance matrix only. Therefore a quality measure is presented that incorporates pixel correlations with the help of the full covariance matrix which allows to differentiate a real structure from noise amplifications. Since magnitude images are often used for diagnosis, also the case of computing the covariance matrix for correlated magnitude pixels from a multidimensional Rice distribution is explored. This is a generalization to the (uncorrelated) pixel by pixel approach and can be used to justify simplifications.

Another approach to compute the covariance matrix of the reconstructed images is to sample from their probability distribution directly. The optimization problem is expressed in the Bayesian sense in terms of likelihood function and prior which is equivalent to the original problem. Then a Metropolis Hastings sampler can be used to draw samples from the posterior distribution. The reconstructed image appears to be the mean of the samples and additionally the covariance matrix can be extracted. This inversion based on MCMC methods is an

independent reconstruction method for MRI that inherently provides a quality measure.

2 Theory of image reconstruction in MRI

2.1 Theory of MRI

MRI enables to reconstruct the internal structure of an object in terms of its nuclear spin density by creating a macroscopic signal from a large number of atoms. The microscopic basis of this resulting signal is the energy splitting of a spin system in an external magnetic field. This splitting can be used to create a spatially variable resonance frequency by applying stationary magnetic field gradients in all three spatial directions. With the help of special excitation pulses a macroscopic magnetization can be induced in a defined region within the object. This macroscopic signal can be detected which makes it possible to reconstruct tomographic images of the nuclear spin density of the sample. Therefore the whole process provides a visualization of the internal structure in the end. Since the wavelength of the occurring radiation is in the radio frequency range MRI is a preferable choice compared to techniques using ionizing radiation.

2.1.1 A spin-$\frac{1}{2}$ particle in a magnetic field

The simplest element with total nuclear spin quantum number $I = \frac{1}{2}$ is the Hydrogen atom which is also one of the main elements in biological tissue (others are ^{13}C, ^{19}F, ^{31}P). Since angular momentum is quantized (with respect to a given z-direction) the expectation value of the z-component of the spin operator $\hat{\mathbf{S}}$ can only take $2I + 1$ discrete values. The total nuclear spin quantum number of the Hydrogen nucleus is $I = s = \frac{1}{2}$, so in total 2 different orientations are expected. Expressing this geometrical interpretation in mathematical terms, the z-component of the spin operator $\hat{S}_z$ has eigenvalues $m\hbar$, where m can take the values $-s, -s+1, ..., s-1, s$. Then the 2 eigenvalues are $\pm\frac{\hbar}{2}$ [5].

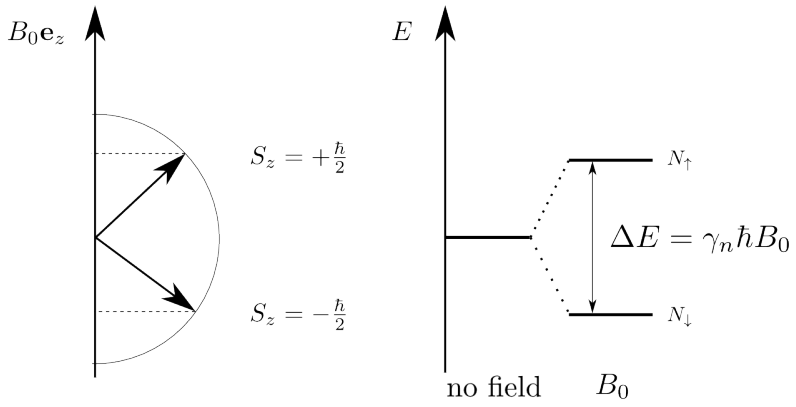


Figure 2: Zeeman effect: Quantization of angular momentum and energy level splitting of a spin $\frac{1}{2}$ particle in a magnetic field.

The behavior of a spin-$\frac{1}{2}$ particle in a magnetic field is described by its magnetic moment operator $\hat{\boldsymbol{\mu}} = \gamma_n \hat{\mathbf{S}}$. The eigenvalues of the z-component of the nuclear magnetic moment is given by $\mu_z = \pm\gamma_n \frac{\hbar}{2}$ where γ_n is the gyromagnetic ratio of the nucleus given by $\gamma_n = g_n \frac{e}{2m_p}$.

The subscript p means proton because it carries the spin and g_n is the g-factor of the nucleus. For the Hydrogen nucleus the gyromagnetic ratio is $\gamma_n = 2.675\ 10^8\ rad\ s^{-1}\ T^{-1}$.

Since there are two different states a spin-$\frac{1}{2}$ particle can occupy, in the presence of an external magnetic field there are also two energetic states. The potential energy of a magnetic moment subject to an external magnetic field is classically $E = -\boldsymbol{\mu}\mathbf{B}_0$. If the magnetic field is directed in z-direction like $\mathbf{B_0} = B_0\mathbf{e_z}$, the energy becomes $E = -\mu_z B_0$. According to the eigenvalues there are two possible energies $E = \pm\gamma_n \frac{\hbar}{2} B_0$ for a spin-$\frac{1}{2}$ particle in an external magnetic field. This is called Zeeman effect. The energy difference between those two states is $\Delta E = \gamma_n \hbar B_0$ corresponding to the Larmor frequency ω_0 which can be interpreted as a precession frequency of a spin around the z-axis. The occupation numbers of these energy levels according to the Boltzmann fraction are given by $\frac{N_\uparrow}{N_\downarrow} = \exp\big(-\frac{\Delta E}{k_b T}\big)$ [18].

2.1.2 Bloch equations - The motion of a spin system

The classical equation of motion for a rotating object in an external field $\mathbf{F}$ in terms of torque $\mathbf{T}$ and angular momentum $\mathbf{L}$ reads $\mathbf{T} = \frac{d\mathbf{L}}{dt} = \mathbf{r} \times \mathbf{F}$. Applying this general equation to the special case of a magnetic moment $\boldsymbol{\mu}$ in an external magnetic field $\mathbf{B}$ yields:

$$\frac{d\mathbf{L}}{dt} = \boldsymbol{\mu} \times \mathbf{B} \tag{2.1}$$

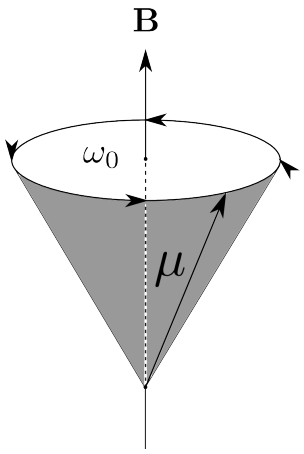


Figure 3: Precession of the magnetic moment around a magnetic field with Larmor frequency ω_0.

Since we want to describe nucleons, we need to find a proper quantum mechanical representation. In the Heisenberg picture this task can be achieved by taking the classical equation of motion for precession and replacing variables with the averages if the according operators. The angular momentum $\mathbf{L}$ will be mapped to the angular momentum operator of the nucleus which is the spin operator $\hat{\mathbf{S}}$ in this case. It can be written in terms of magnetic moments as $\hat{\mathbf{S}} = \frac{1}{\gamma_n}\hat{\boldsymbol{\mu}}$, where $\gamma_n = g_n \frac{e}{2m_p}$ is a combination of physical constants. Multiplying equation (2.1) with γ_n on both sides gives:

$$\frac{d\langle\hat{\boldsymbol{\mu}}\rangle}{dt} = \gamma_n \langle\hat{\boldsymbol{\mu}}\rangle \times \langle\hat{\mathbf{B}}\rangle \tag{2.2}$$

To describe the macroscopic magnetization, the the sum over the expectation values of the magnetic moment of all N microscopic spins is taken: $\mathbf{M} = \frac{1}{N}\sum_{i=1}^{N}\langle\boldsymbol{\mu}_i\rangle$. Now the equation can be split into a transverse and a longitudinal part (in z-direction). In the absence of relaxation processes (**non-interacting spins**) we get

$$\frac{d\mathbf{M}_\perp}{dt} = \gamma_n \mathbf{M}_\perp \times \mathbf{B}$$
$$\frac{dM_z}{dt} = 0$$

The above equation is valid for a magnetization formed by a large number of non-interacting spins which is equivalent (up to a factor) to a single spin. Usually the considered spins are interacting with each other such that the system will go back to equilibrium after an external excitation and the equations get damping factors due to spin-spin and spin-lattice interactions. If we identify spin-spin interactions as a process affecting the transverse part only and spin-lattice interactions affecting the longitudinal part only and we model those processes as linear relaxations with (spatially dependent) relaxation times T_1 and T_2, we obtain the **Bloch equations for interacting spins**.

$$\boxed{\begin{aligned}\frac{d\mathbf{M}_\perp}{dt} &= \gamma \mathbf{M}_\perp \times \mathbf{B} - \frac{\mathbf{M}_\perp}{T_2} \\ \frac{dM_z}{dt} &= -\frac{M_z - M_{z,0}}{T_1}\end{aligned}} \tag{2.3}$$

They describe the motion of the macroscopic magnetization formed by the ensemble of microscopic nuclear spins back to equilibrium after an excitation pulse in the presence of an external magnetic field.

2.1.3 Spatial excitation and detection of the induced magnetization

In the following a short summary of the most essential steps in the derivation of the equation that describes MRI signal acquisition is shown. It mainly follows the presentation given in the book *Magnetic Resonance Imaging: Physical Principles and Sequence Design* by Haacke, Brown, Thompson and Venkatesan [3].

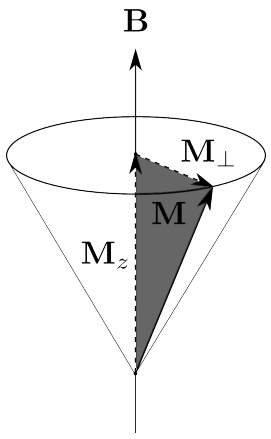


Figure 4: Transverse ($\mathbf{M}_\perp$) and longitudinal part ($\mathbf{M_z}$) of the macroscopic magnetization vector ($\mathbf{M}$) with excitation pulse $\mathbf{B_{exc}}$ in the transverse plane.

A spin ensemble in the external magnetic field will not generate any measurable signal without transitions from between the lower and the upper level. So the spins from the lower level need to be pumped to the upper level and then their their relaxation can be measured. Another difficulty is that the total transverse magnetization cancels to zero since the phases of the spins are distributed randomly. Therefore the spins do not only have to be excited but also they have to have the same phase. In practical applications the spins are excited by switching on a short external EM-excitation pulse $\mathbf{B_{exc}} = B_1(\cos(\omega_0 t)\mathbf{e_x} - \sin(\omega_0 t)\mathbf{e_y})$ in the transverse plane. This will generate a net magnetization in the transverse plane by tilting all spins into this plane having the same phase. As soon as the external pulse is switched off again, the spin system moves back to equilibrium and the transverse net magnetization decays according to the solutions of the Bloch equations.

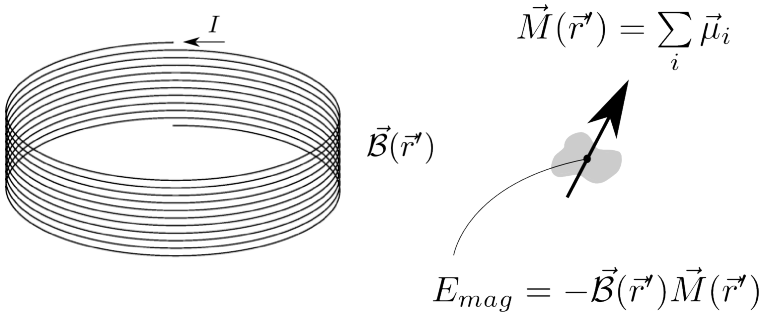


Figure 5: Principle MRI-setup: A coil loaded with a certain load I corresponds to a magnetic field $\mathcal{B}(\mathbf{r}')$ which is interacting with the magnetization $\mathbf{M}(\mathbf{r}')$ of a sample. A decay of the induced magnetization in the sample according to the Bloch equations leads to an (additional) voltage in the coil according to Faraday's law. The quantity that is measured in MRI is the magnetic energy at a point $\mathbf{r}'$, formed by the coil's field and the magnetization. This can also be interpreted as the principle of reciprocity (see 2.1.3).

The detection of the macroscopic magnetization can be measured in a receiver coil. According to Faraday's law of induction, a temporal change of magnetic flux (due to the decay of magnetization) creates an electromotive force (voltage) EMF $\propto -\frac{d}{dt}\Phi(t)$. Magnetic flux is defined as $\Phi(t) = \int_{\text{coilarea}} \mathbf{B}d\mathbf{f}$, where $\mathbf{B} = \nabla \times \mathbf{A}$. With Stokes theorem that can be rewritten as $\Phi(t) = \oint_{wire} \mathbf{A}d\mathbf{l}$. For Coulomb gauge and quasi-stationary fields and with the contribution of the magnetization to the current density $\mathbf{j}_M(\mathbf{r},t) = \nabla_{\mathbf{r}} \times \mathbf{M}(\mathbf{r},t)$, the vector potential reads

$$\mathbf{A}(\mathbf{r},t) = \frac{\mu_0}{4\pi}\int_V \frac{\mathbf{j}_M(\mathbf{r}',t)}{\|\mathbf{r}-\mathbf{r}'\|}d^3\mathbf{r}' = \frac{\mu_0}{4\pi}\int_V \frac{\nabla_{\mathbf{r}'} \times \mathbf{M}(\mathbf{r}',t)}{\|\mathbf{r}-\mathbf{r}'\|}d^3\mathbf{r}'$$

The magnetic flux is defined as the integral of the vector potential over the wire

$$\Phi(t) = \oint_{wire} \mathbf{A}d\mathbf{l} = \oint_{wire} \frac{\mu_0}{4\pi}\int_V \frac{\nabla_{\mathbf{r}'} \times \mathbf{M}(\mathbf{r}',t)}{\|\mathbf{r}-\mathbf{r}'\|}d^3\mathbf{r}'d\mathbf{l} = \int_V \frac{\mu_0}{4\pi}\oint_{wire} \frac{\nabla_{\mathbf{r}'} \times \mathbf{M}(\mathbf{r}',t)}{\|\mathbf{r}-\mathbf{r}'\|}d\mathbf{l}d^3\mathbf{r}'$$

$$= \int_V \mathbf{M}(\mathbf{r}',t)\underbrace{\frac{\mu_0}{4\pi}\oint_{wire} \frac{d\mathbf{l}\times(\mathbf{r}-\mathbf{r}')}{\|\mathbf{r}-\mathbf{r}'\|^3}d^3\mathbf{r}'}_{\mathcal{B}(\mathbf{r}')} = \int_V \mathbf{M}(\mathbf{r}',t)\mathcal{B}(\mathbf{r}')d^3\mathbf{r}' \tag{2.4}$$

With the help of some vector identities the surface integral was reformulated such that $\mathcal{B}(\mathbf{r}')$

represents **Biot-Savart's law** for a constant current $I = 1$:

$$\mathcal{B}(\mathbf{r}') = \frac{\mu_0}{4\pi} \oint_{wire} \frac{Id\mathbf{l} \times (\mathbf{r} - \mathbf{r}')}{\|\mathbf{r} - \mathbf{r}'\|^3} \tag{2.5}$$

Replacing $\mathbf{M}$ in equation (2.4) with the solutions of the Bloch equations and pulling and executing the negative time-derivative for the EMF inside the integral yields the time dependent signal in the receiver coil. Taking into account that the Larmor frequency ω_0 is orders of magnitude larger than $\frac{1}{T_1}$ and $\frac{1}{T_2}$, a simplified solution can be derived where the z-component of $\mathbf{M}$ is zero. The remaining components M_x and M_y can be expressed as real and imaginary part of a complex quantity M_+ whose magnitude is $M_\perp$. Depending on the RF-pulse, $\phi_0(\mathbf{r}')$ is the initial phase of the spins at a certain position. With the help of a projection parameter $\theta_\mathcal{B}(\mathbf{r}')$ the magnetic field becomes $\mathcal{B}_\perp(\mathbf{r}')$ and after using some trigonometric formulas the EMF becomes

$$EMF \propto \omega_0 \int_V M_\perp(\mathbf{r}', t)\mathcal{B}_\perp(\mathbf{r}')\sin\big(\omega_0 t + \theta_\mathcal{B}(\mathbf{r}') - \phi_0(\mathbf{r}')\big)d^3\mathbf{r}'$$

Demodulation and representation as a complex signal

The EMF is measured as a voltage in a receiver coil. It represents the macroscopic transverse magnetization vector $\mathbf{M}_\perp$ of the given sample. It is convenient to assemble the x-component of the vector as $\mathrm{Re}(s(t))$ and the y-component as $\mathrm{Im}(s(t))$ of a complex valued signal. This can be obtained by performing a quadrature amplitude demodulation (QAD), where the signal is split into two equivalent ones which are multiplied by $\sin(\omega_0 t + \delta\omega t)$ and $\cos(\omega_0 t + \delta\omega t)$ respectively. $\delta\omega$ is a small frequency offset from the Larmor frequency ω_0 and $\Omega = \omega_0 + \delta\omega$ is the demodulation frequency. After passing them through a low-pass filter, the two signals are reassembled again as a complex signal $s(t) = s_{Re} + is_{Im}$ which has the form

$$s(t) \propto \omega_0 \int_V M_\perp(\mathbf{r}', 0)\mathcal{B}_\perp(\mathbf{r}')e^{i((\Omega-\omega_0)t + \phi_0(\mathbf{r}') - \theta_\mathcal{B}(\mathbf{r}'))}d^3\mathbf{r}'$$

Above's magnetization magnitude is time-independent since the exponential T_1 and T_2-relaxations have been neglected. The magnitude of the transverse magnetization is proportional to the spin density one is interested in. When the static magnetic field $\mathbf{B}_0$ (which corresponds to a Larmor frequency ω_0) is sufficiently homogeneous, $\theta_\mathcal{B}(\mathbf{r}')$ and $\phi_0(\mathbf{r}')$ are spatial independent and can - as well as the proportionality constant - merged with $M_\perp(\mathbf{r}', 0)$ to form the spin density $\rho(\mathbf{r}')$. That procedure leads to the signal equation for one distinct Larmor frequency ω_0 and one demodulation frequency Ω.

$$s(t) = \int_V \rho(\mathbf{r}')\mathcal{B}_\perp(\mathbf{r}')e^{i((\Omega-\omega_0)t)}d^3\mathbf{r}' \tag{2.6}$$

Coil sensitivities and the principle of reciprocity

The expression $\mathcal{B}_\perp(\mathbf{r}')$ can be interpreted as the coil sensitivity of the receiver coil and will be denoted as $c_{3D}(\mathbf{r}')$ in the following. Given a defined current in the coil (coil load) and a certain coil geometry, a spatial dependent magnetic field is created according to Biot-Savart's law according to equation (2.5). On the other hand a given magnetic field will also create a defined current in the coil. This is called principle of reciprocity or antenna theorem [22]. It basically states that the sensitivity of a transmitting RF-unit at a point $\mathbf{r}'$ is equal to the sensitivity of the same arrangement used as a receiving unit. In equation (2.4) this principle can be observed: the MRI-signal is formed by an integral over magnetic field energies, it does not matter how these energies are formed (be it by the classical MRI-setup using the coils as receivers or the other way around). That insight can be useful by calculating the coil sensitivities $c_{3D}(\mathbf{r}')$, which need to be known for SENSE image reconstruction.

2.1.4 The signal equation

For spatial encoding of the sample volume **B**-field gradients are switched on in the x-, y- and z-direction for a certain time period. The Gradients are defined as $G_i(t) := \frac{\partial B_i(t)}{\partial x_i}$ for $i \in x, y, z$ such that the magnetic field that encodes in z-direction will be $B_z(z,t) = B_0 + zG_z(t)$. That means that the Larmor frequency will be a function of space and time now. For the z-direction only this denotes as $\omega(\mathbf{r},t) = \omega_0 + \omega_G(\mathbf{r},t) = \omega_0 + \gamma z G_z(t)$. If we define a phase $\phi(\mathbf{r},t) = -\int_0^t \omega(\mathbf{r},t')dt'$ and set the demodulation frequency $\Omega := \omega_0$ (with no offset $\delta\omega$) the exponent in equation (2.6) is rewritten as

$$s(t) = \int_V \rho(\mathbf{r}) c_{3D}(\mathbf{r}) e^{-i\gamma z \int_0^t G_z(t')dt'} d^3\mathbf{r}'$$

For 3D-encoding 3 gradients are needed which extends the exponent by 2 more integrals. With a redefinition of these integrals as k-vectors $k_i(t) = \frac{\gamma}{2\pi}\int_0^t G_i(t')dt'$ the full 3D-encoding MRI signal equation can be written down. It turns out to be the Fourier-transform of the product of the spin density with the coil sensitivity:

$$s(\mathbf{k}(t)) = \int_V \rho_{3D}(\mathbf{r}) c_{3D}(\mathbf{r}) e^{-i2\pi\mathbf{k}(t)\mathbf{r}} d^3\mathbf{r}'$$

For simplification reasons only 2-dimensional signals will be considered in the following. That means that a certain slice in the z-direction of the scanned object has been fixed and ρ and c are reduced to 2D-functions now. Integrating over the z-direction yields

$$s(k_x(t), k_y(t)) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \rho(x,y) c(x,y) e^{-i2\pi(k_x(t)x + k_y(t)y)} dxdy$$

For the sake of a more convenient notation this equation will be rewritten once more as the 2-dimensional signal equation for a single coil.

$$s(\mathbf{k}) = \int_{Slice} \rho(\mathbf{r}) c(\mathbf{r}) e^{-i2\pi\mathbf{kr}} dxdy \tag{2.7}$$

omitting the time dependence of $\mathbf{k}$ and keeping in mind that its z-component is constant which can be absorbed in ρ because of proportionality. Care has to be taken that the spin density $\rho(\mathbf{r})$ and the coil sensitivity $c(\mathbf{r})$ are complex numbers.

2.1.5 The signal equation for parallel MRI (PRMI)

In PMRI multiple receiver coils are used which are spatially distributed around the scanned object to measure the signal parallely. Each coil will record the signal from another direction and hence has a certain coil sensitivity $c_j(x,y)$ in image-space. The coil sensitivity for coil j is a map $c_j(x,y)$ from $\mathrm{R} \times \mathrm{R} \to \mathrm{C}$ because the spin density is a complex signal. The coil sensitivity represents the spatial resolution of the receiver coil j in the tomograph in image space as a complex number. Equation (2.7) is valid for each coil j and we can write

$$\boxed{s_j(\mathbf{k}) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \rho(\mathbf{r}) c_j(\mathbf{r}) e^{-i2\pi\mathbf{kr}} dxdy} \tag{2.8}$$

2.2 Numerical Implementation of parallel MRI: The forward operator

The numerical implementation of the model for parallel MRI is given by the forward operator. For SENSE (compare the original paper of Pruessmann [19], [14] and [2]) it consists of (1) the Fourier transform, (2) the coil sensitivities and (3) the sampling trajectory which defines the undersampling pattern. In principle all these parts could be given as matrices but matrix sizes of 128 x 128 and more prohibit that due to memory problems. That is why in the implementation used for this thesis the forward operator is given by a function with two subarguments. The Fourier transform is given by a FFT whereas the sampling trajectory, the coil sensitivities, the argument and the target of the operator are given by matrices. That has the drawback that for operations that need the matrix representation of the forward operator (like condition number and eigenvalues) the matrix has to be created first from the canonical base vectors. Still speed and memory arguments are more convincing. The forward operator represents the model relating measurements with the underlying physical process. In the case of MRI this is the Fourier transform according to equation (2.7). For discretization purposes we rewrite this equation and interpret the signal $s_j(t)$ as a data vector $\mathbf{y}_j$, where for Cartesian sampling the elements in this vector are ordered starting with the signal coming from the upper left part of k-space going down in the same column and then starting again in the second column with the first element and so on. The spin density is the actual image of interest, so we will store its values in a vector $\mathbf{x}$ where the elements are ordered in a similar fashion. Accordingly we will reorder the values of the coil sensitivities and write them as a vector $\mathbf{c}_j$.

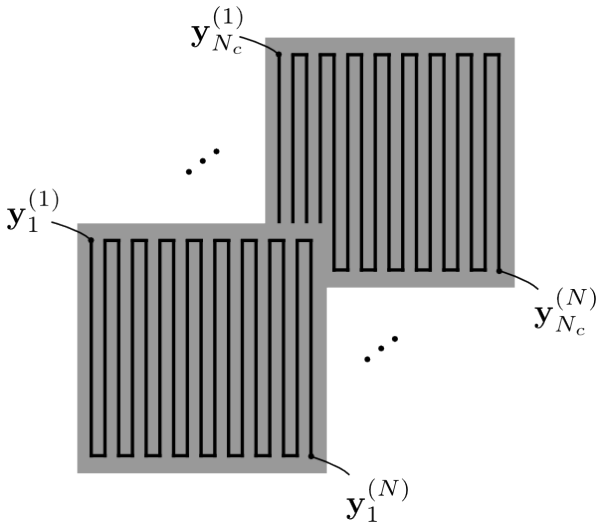


Figure 6: Arrangement scheme of k-space data.

After the introduction of the 2D discrete Fourier transform - which is written as a matrix $\mathcal{F}$ - a discrete version of equation (2.7) is obtained.

$$\mathbf{y}_j = \mathcal{F}\big(\mathbf{c}_j \circ \mathbf{x}\big)$$

The operator $\circ$ depicts the Hadamard-product which is an elementwise multiplication of the elements of two vectors (or matrices) of the same size. In the operator notation this writes

$$\mathbf{y}_j = \mathcal{A}_j \mathbf{x}$$

where $\mathcal{A}_j$ is a Fourier matrix that matrix-multiplies the Hadamard product of the coil sensitivities with the argument. The coil sensitivities are assumed to be known, e.g. they can be approximated before or during the scan.

$$\mathcal{A}_j := \mathcal{F}\mathbf{c}_j\circ$$

In parallel MRI multiple coils are used to increase SNR or to obtain more information at the same time. Then the image $\mathbf{x}$ is composed of all the coils' information. Therefore it is necessary to assemble a data vector $\mathbf{y}$ with N_c components $\mathbf{y}_j$. The operator R generates N_c copies of $\mathbf{x}$ and is a matrix with N_c rows formed by unity matrices.

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_{N_c} \end{pmatrix} = \begin{pmatrix} \mathcal{F}\big(\mathbf{c}_1 \circ \mathbf{x}\big) \\ \vdots \\ \mathcal{F}\big(\mathbf{c}_{N_c} \circ \mathbf{x}\big) \end{pmatrix} = \begin{pmatrix} \mathcal{A}_1\mathbf{x} \\ \vdots \\ \mathcal{A}_{N_c}\mathbf{x} \end{pmatrix} = \underbrace{\begin{pmatrix} \mathcal{A}_1 \\ \vdots \\ \mathcal{A}_{N_c} \end{pmatrix}}_{\mathcal{A}} R\,\mathbf{x}$$

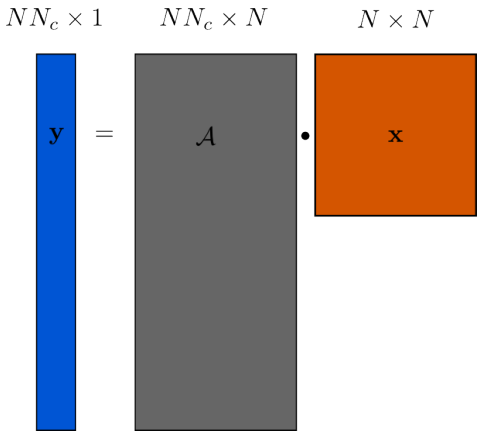


Figure 7: Graphical representation of the forward problem $\mathbf{y} = \mathcal{A}\mathbf{x}$.

The (generalized) forward operator $\mathcal{A}$ for multiple coils has the following shape:

$$\mathcal{A} = \begin{pmatrix} \mathcal{A}_1 \\ \vdots \\ \mathcal{A}_{N_c} \end{pmatrix} R = \begin{pmatrix} \mathcal{A}_1 \\ \vdots \\ \mathcal{A}_{N_c} \end{pmatrix} \begin{pmatrix} \mathbf{1} \\ \vdots \\ \mathbf{1} \end{pmatrix} = \begin{pmatrix} \mathcal{F}\mathbf{c}_1\circ \\ \vdots \\ \mathcal{F}\mathbf{c}_{N_c}\circ \end{pmatrix} \begin{pmatrix} \mathbf{1} \\ \vdots \\ \mathbf{1} \end{pmatrix}$$

For the CG-SENSE inversion also the Hermitian conjugate of the forward operator is needed. The complex conjugate transpose operator of the forward operator is composed of the matrix of the inverse Fourier transform $\mathcal{F}^{-1}$ and the complex conjugate of the coil sensitivities.

$$\begin{aligned} \mathcal{A}^\dagger &= \begin{pmatrix} \mathbf{1} \\ \vdots \\ \mathbf{1} \end{pmatrix}^\dagger \begin{pmatrix} \mathcal{A}_1 \\ \vdots \\ \mathcal{A}_{N_c} \end{pmatrix}^\dagger = \begin{pmatrix} \mathbf{1} & \dots & \mathbf{1} \end{pmatrix} \begin{pmatrix} \mathcal{A}_1^\dagger \\ \vdots \\ \mathcal{A}_{N_c}^\dagger \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{1} & \dots & \mathbf{1} \end{pmatrix} \begin{pmatrix} \big(\mathcal{F}\mathbf{c}_1 \circ\big)^\dagger \\ \vdots \\ \big(\mathcal{F}\mathbf{c}_{N_c} \circ\big)^\dagger \end{pmatrix} = \begin{pmatrix} \mathbf{1} & \dots & \mathbf{1} \end{pmatrix} \begin{pmatrix} \mathbf{c}_1^* \circ \mathcal{F}^{-1} \\ \vdots \\ \mathbf{c}_{N_c}^* \circ \mathcal{F}^{-1} \end{pmatrix} \end{aligned}$$

Applied to an argument $\mathbf{y}$ with N_c components formed by row-vectors $\mathbf{y}_j$ corresponding to the data collected by each coil (and using a replicating operator R consisting of unity matrices) this can be rewritten as

$$\mathcal{A}^\dagger \mathbf{y} = \sum_{j=1}^{N_c} \mathbf{c}_j^* \circ \mathcal{F}^{-1} \mathbf{y}_j \tag{2.9}$$

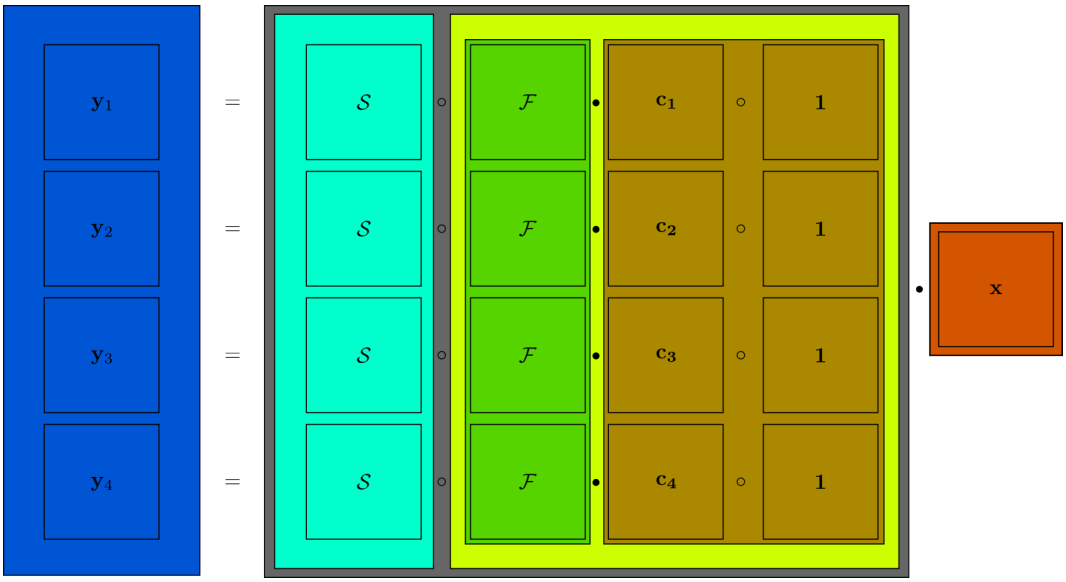


Figure 8: Graphical structure of the forward problem for parallel MRI with k-space data of the individual coils $\mathbf{y}_1, \ldots, \mathbf{y}_4$, the sampling operator $\mathcal{S}$ (which generates the k-space trajectory), a 2D-DFT-matrix $\mathcal{F}$, the individual coil sensitivities $\mathbf{c}_1, \ldots, \mathbf{c}_4$ and the image $\mathbf{x}$ (which is replicated by R).

2.2.1 Sampling trajectories

A sampling trajectory is basically the sampling pattern in k-space. In this thesis only Cartesian sampling is considered, whilst others like radial, spiral or random sampling exist. The following figure shows sampling patterns for different acceleration factors where whole k-space lines have been left out where the k-space center remains fully sampled.

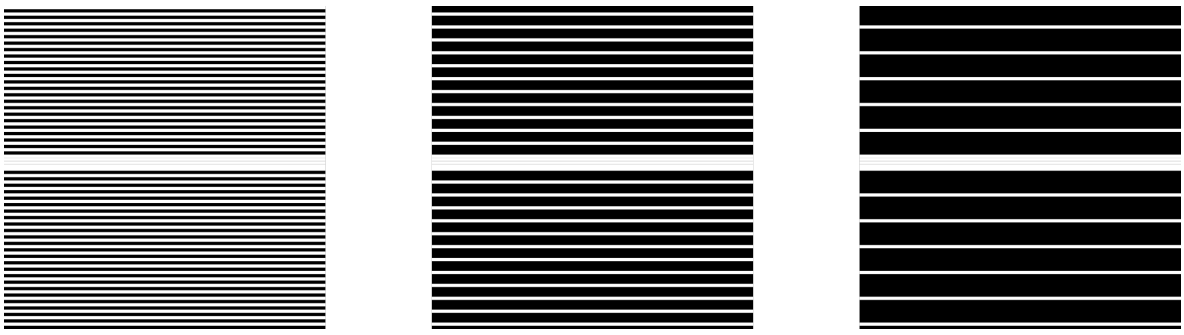
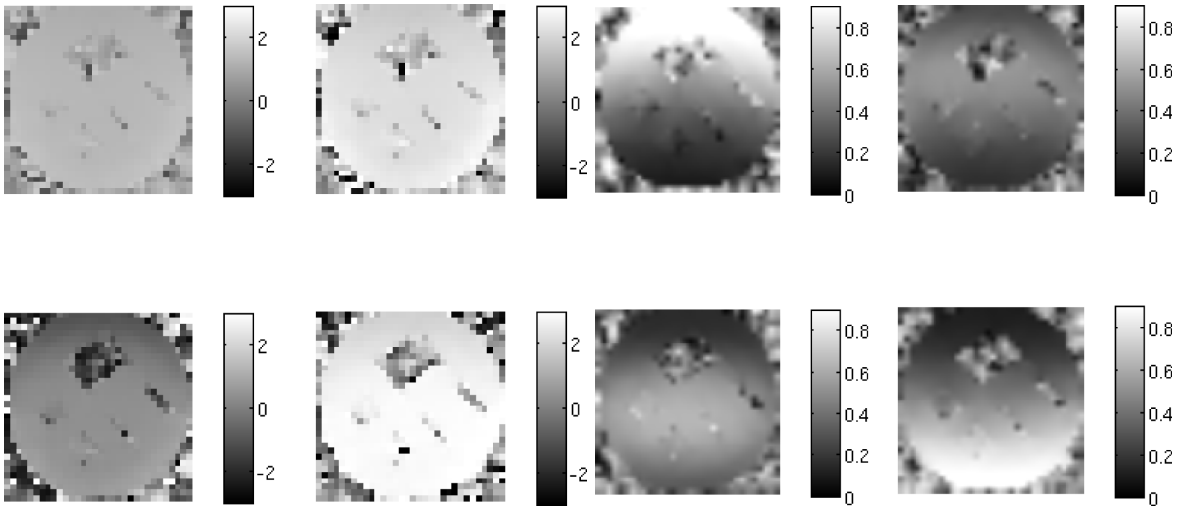


Figure 9: Sampling trajectories for different acceleration factors: acc = 2 (left), acc = 4 (middle), acc = 8 (right).

2.2.2 Coil sensitivities

For multiple detector coils in Parallel MRI (PMRI) the coil sensitivities can be determined experimentally from the scanner by estimating them in image space (SENSE, compare [19]). This is done in a preliminary measurement. That is why in the following figure the structure of the measured phantom is visible - though not very precisely because only the body coil of the scanner is used.



(a) Phase of typical coil sensitivities. (b) Magnitude of typical coil sensitivities.

2.2.3 Decorrelation using the Cholesky transform

The weighted LSQ problem in equation 3.10 uses the covariance matrix of the data noise in the cost functional. One practical way to include it in the problem formulation from the beginning is to decorrelate the data and alter the model operator. This is achieved by applying the Cholesky decomposition to the covariance matrix. For every positive-definite and hermitian matrix A exists a Cholesky decomposition as

$$A = L^\dagger L$$

Applying this to $A := C^{-1}$ we get for the data fidelity term in the cost functional

$$\begin{aligned}(\mathbf{y} - \mathcal{A}\mathbf{x})^\dagger C^{-1}(\mathbf{y} - \mathcal{A}\mathbf{x}) &= (\mathbf{y} - \mathcal{A}\mathbf{x})^\dagger L^\dagger L(\mathbf{y} - \mathcal{A}\mathbf{x}) \\ &= (L\mathbf{y} - L\mathcal{A}\mathbf{x})^\dagger (L\mathbf{y} - L\mathcal{A}\mathbf{x}) \\ &= (\tilde{\mathbf{y}} - \tilde{\mathcal{A}}\mathbf{x})^\dagger (\tilde{\mathbf{y}} - \tilde{\mathcal{A}}\mathbf{x})\end{aligned}$$

Now the problem is rewritten as uncorrelated data $\tilde{\mathbf{y}} := L\mathbf{y}$ and with an altered forward operator $\tilde{\mathcal{A}} := L\mathcal{A}$. The transform can be done after acquiring data at input level of any further algorithms and leads to uncorrelated noise with unity covariance. It can also be seen as a preconditioner of the problem that will increase convergence. Especially for MCMC methods that can increase speed since in the symmetric case of uncorrelated components the chain can move into the region of interest with one big step instead of many small ones.

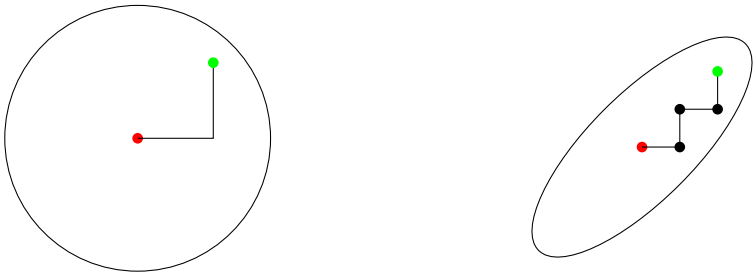


Figure 11: Correlated noise covariance (Left): The Markov chain moves slowly due to asymmetric region of acceptance. Uncorrelated noise covariance (Right): The Markov chain moves in big steps due to symmetric region of acceptance.

3 Inverse Problems

3.1 Introduction and definitions

Usually certain measured data like the gravitational field of the earth is linked to a physical quantity like the earth's mass density with the help of a physical model - in this case Newton's gravitational law. Mathematically speaking we consider a physical model represented as a matrix $\mathcal{A} \in \mathbf{C}^{m \times n}$ with n model parameters assembled in a vector $\mathbf{x} \in \mathbf{C}^{n \times 1}$ and m given data points assembled to a vector $\mathbf{y} \in \mathbf{C}^{m \times 1}$.

$$\mathbf{y} = \mathcal{A}\mathbf{x}$$

We have a linear model operator $\mathcal{A}$ which lets us calculate the data $\mathbf{y}$ given the model parameters $\mathbf{x}$ as input. This is called **forward problem**. However, more frequently we are facing the opposite task: We have certain data $\mathbf{y}$ obtained by a measurement and want to determine our model parameters $\mathbf{y}$ according to $\mathcal{A}$. If the measurement could be done without errors and we knew the inverse of our model operator A, it would be easy to accomplish that. Unfortunately in the most cases these conditions are not fulfilled: The measured data is corrupted by a certain noise $\mathbf{n}$. Additionally the model operator A might be ill-conditioned such that we cannot simply invert it. So the complexity of finding $\mathbf{x}$ increases. Denoted by the following equation 3.1 an **inverse problem** is defined as: "Given the measurement $\mathbf{y}$ with noise $\mathbf{n}$, find $\mathbf{x}$ such that $\mathbf{y} = A\mathbf{x} + \mathbf{n}$ is fulfilled". [21].

$$\boxed{\mathbf{y} = A\mathbf{x} + \mathbf{n}} \tag{3.1}$$

For solving the inverse problem the solutions $\mathbf{x}$ of equation (3.1) have to be found. Depending on the condition of the matrix $\mathcal{A}$ and therefore on the ill-posedness of the problem, it has either a unique solution, many solutions or no solution at all. A definition of an ill-posed problem has been given by Jaques Hadamard in 1902 [11].

A problem is called **ill-posed problem** when it meets either one of the following criteria:

1. the solution does not exist
2. the solution is not unique
3. the solution is extremely sensitive with respect to initial data

In the following some general tools for analyzing and solving (more or less) ill-posed problems are briefly discussed, mostly following the books of Per Christian Hansen [12] and David G. Luenberger [16].

3.1.1 The Moore-Penrose pseudo-inverse

The Moore-Penrose pseudo-inverse is the generalized inverse of a matrix. For any matrix $\mathcal{A} \in \mathbf{C}^{m\times n}$ there exists a unique matrix $\mathcal{A}^+ \in \mathbf{C}^{n\times m}$ - the Moore-Penrose pseudo-inverse - with the following properties:

1. $\left(\mathcal{A}\mathcal{A}^+\right)^\dagger = \mathcal{A}\mathcal{A}^+$
2. $\left(\mathcal{A}^+\mathcal{A}\right)^\dagger = \mathcal{A}^+\mathcal{A}$
3. $\mathcal{A}\mathcal{A}^+\mathcal{A} = \mathcal{A}$
4. $\mathcal{A}^+\mathcal{A}\mathcal{A}^+ = \mathcal{A}^+$

From the first two properties follows that $\mathcal{A}\mathcal{A}^+$ and $\mathcal{A}^+\mathcal{A}$ are hermitian, from the last two it follows that $\mathcal{A}\mathcal{A}^+$ is the identity map for column vectors (which is not necessarily the identity matrix $\mathbf{1}$). Only if $\mathcal{A}$ has full rank $\mathcal{A}\mathcal{A}^+$ is the identity matrix $\mathbf{1}$. Then $\mathcal{A}$ is invertable and the Moore-Penrose pseudo-inverse is identical with the inverse matrix $\mathcal{A}^{-1}$. Furthermore for full rank matrices the condition $\left(AB\right)^+ = B^+A^+$ holds.

3.1.2 The singular value decomposition (SVD)

The singular value decomposition works for general matrices. As an example we consider a rectangular matrix $\mathcal{A} \in \mathrm{C}^{m\times n}$. Without loss of generality for $m \geq n$ there exists a decomposition of $\mathcal{A}$ such that

$$\mathcal{A} = U\Sigma V^\dagger = \sum_{i=1}^{\mathrm{rank}(\mathcal{A})} \mathbf{u}_i \sigma_i \mathbf{v}_i^\dagger$$

For matrices with $m < n$ this decomposition can be applied to the complex conjugate transpose of $\mathcal{A}$ in the same manner. The matrices $U \in \mathrm{R}^{m\times m}$ and $V \in \mathrm{R}^{n\times n}$ are always full-rank matrices assembled by the left-singular vectors $(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m)$ and the right-singular vectors $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$. The columns of each U and V are orthogonal and the matrices obey the property $U^\dagger U = \mathbf{1}_m$ and $V^\dagger V = \mathbf{1}_n$. The matrix $\Sigma \in \mathrm{R}^{m\times n}$ is a diagonal matrix of the real singular values of $\mathcal{A}$ with non-increasing diagonal elements $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$. The result of the SVD can be used to compute the pseudo-inverse. For this purpose the following relation is useful:

$$\mathcal{A}^+ = V\Sigma^+ U^\dagger = \sum_{i=1}^{\mathrm{rank}(\mathcal{A})} \mathbf{v}_i \sigma_i^{-1} \mathbf{u}_i^\dagger$$

The pseudo-inverse of Σ is still a real-valued diagonal matrix with the reciprocal elements σ_i^{-1} as diagonal elements. When the matrix has full rank the limit in the sum will become $\min(m, n)$ and $\mathcal{A}^+ = \mathcal{A}^{-1}$.

Furthermore the SVD is an important tool to quantify the ill-conditioning of a matrix. Analyzing the singular values assembled in Σ provides information on the rank-deficiency of the matrix. If there is a bunch of high singular values separated of another region with comparably small ones, it is a strong hint that the matrix contains redundant information and therbefore has low rank. In the case of a smooth decay of the singular values the opposite cannot be deduced and $\mathcal{A}$ might still have low rank.

3.1.3 The condition number of a matrix

With respect to a linear system $\mathbf{y} = \mathcal{A}\mathbf{x}$ the condition number of $\mathcal{A}$ is a measure of how sensitively it maps small changes in the argument $\mathbf{x}$ to the target space. In other words it gives an idea about how much the output vectors $\mathbf{y}$ will vary if the input parameter $\mathbf{x}$ is varied. The condition number is a measure for the third criterion for well-posed problems - next to (1) existence and (2) number of solutions. If the condition number is high, $\mathcal{A}$ is very sensitive to perturbations in $\mathbf{x}$. The condition number is defined with the help of the results of the SVD.

$$\mathrm{cond}(\mathcal{A}) = \|\mathcal{A}\|_2 \|\mathcal{A}^+\|_2 = \frac{\sigma_{\max}}{\sigma_{\min}} \tag{3.2}$$

The notation $\|\bullet\|_2$ is the extension of the L_2-vector norm to matrices. It is defined as

$$\|\mathcal{A}\|_2 := \sup_{\mathbf{x}\neq\mathbf{0}} \left\{ \frac{\|\mathcal{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \right\}$$

Since the relation above is valid for any vector $\mathbf{x}$, the denominator plays the role of a normalization. It can be shown that the supremum corresponds to the square of the largest singular value.

From the first part of equation (3.2) one can see that the condition number of a matrix and of its pseudo-inverse are identical. That plays an important role for the solution of the inverse problem $\mathbf{x} = \mathcal{A}^{-1}\mathbf{y}$ where one's task is to find the parameters $\mathbf{x}$ given certain input data $\mathbf{y}$. If and only if $\mathcal{A}$ has full rank and therefore its inverse is the pseudo-inverse, the sensitivity of the solution with respect to small perturbations in the data is the same for the forward and the backward problem due to the identity of the condition number of $\mathcal{A}$ and $\mathcal{A}^+$. However, in most cases $\mathcal{A}$ is rank-deficient and for getting information about the sensitivity of the solution with respect to small changes in the data (for example due to noise in the data measurement process) when solving the inverse problem one needs to compute the condition number of $\mathcal{A}^{-1}$ or its approximation.

3.2 The Bayesian approach to inverse problems

The task of solving an inverse problem is to find those parameters of a given model which approximate the measured data best, taking into account that the measurement process underlies noise according to a certain distribution. Usually the problem is not (too) well conditioned such that the inversion of matrix $\mathcal{A}$ is not possible. From the Bayesian probability theory's point of view all the information about the problem can be used to construct a probability density function (PDF) for the solution $\mathbf{x}$ (posterior PDF). Its maximum is the most probable solution (MAP-solution) and will therefore solve the problem in the statistical

sense. To construct the posterior PDF one can apply the Bayesian theorem to the inverse problem. Generally, for two propositions X, Y it reads

$$P(Y|X)\ P(X) = P(X|Y)\ P(Y)$$

If we assume that Y and X are based on another proposition Z (for example all the additional information we have about the problem), we can rewrite that as

$$\boxed{P(Y|X,Z)\ P(X|Z) = P(X|Y,Z)\ P(Y|Z)} \tag{3.3}$$

Now we will assign each proposition with a specific meaning according to the problem (compare [4]). It consists of a model $\mathcal{A}$ with its parameters $\mathbf{x}$ to describe data $\mathbf{y}$. Since every physical measurement is biased with a certain uncertainty (thermal noise, finite precision of an ADC, etc.) the data has a certain known noise distribution given by a covariance matrix $\mathcal{C}$. We define:

X	$\longrightarrow\ \mathbf{x}$	n parameters written as $\mathbf{x} = (x_1, x_2, ..., x_{n-1}, x_n)$
Y	$\longrightarrow\ \mathbf{y}$	m data points written as $\mathbf{y} = (y_1, y_2, ..., y_{m-1}, y_m)$
Z	$\longrightarrow\ \mathcal{A}, \mathcal{C}$	mathematical model and information about noise in the data

All the information about the character of the measurement process is included in above's propositions. Using equation (3.3) for probability densities yields

$$\underbrace{p(\mathbf{y}|\mathbf{x},\mathcal{A},\mathcal{C})}_{\text{Likelihood function}}\ \underbrace{p(\mathbf{x}|\mathcal{A},\mathcal{C})}_{\text{Prior distribution}} = \underbrace{p(\mathbf{x}|\mathbf{y},\mathcal{A},\mathcal{C})}_{\text{Posterior distribution}}\ \underbrace{p(\mathbf{y}|\mathcal{A},\mathcal{C})}_{\text{Evidence (constant)}}$$

As mentioned above, the PDF we are interested in for solving the inverse problem is the posterior PDF. It can be extracted from equation 3.4 easily. The evidence term does not depend on the parameters and can be treated as a proportionality for the posterior.

$$p(\mathbf{x}|\mathbf{y},\mathcal{A},\mathcal{C}) = \frac{p(\mathbf{y}|\mathbf{x},\mathcal{A},\mathcal{C})\ p(\mathbf{x}|\mathcal{A},\mathcal{C})}{p(\mathbf{y}|\mathcal{A},\mathcal{C})} \propto p(\mathbf{y}|\mathbf{x},\mathcal{A},\mathcal{C})p(\mathbf{x}) \tag{3.4}$$

Up to now we are using general PDFs without having specified a certain type of function for any PDF. The prior PDF can be set equal to one if we do not have any a priori information about the structure of the solutions. In this case the posterior PDF looks like

$$p(\mathbf{y}|\mathbf{x},\mathcal{A},\mathcal{C}) = e^{-\frac{1}{2}\left(\mathbf{y}-\mathcal{A}\mathbf{x}\right)^\dagger C^{-1}\left(\mathbf{y}-\mathcal{A}\mathbf{x}\right)} \tag{3.5}$$

The MAP solution in this case is called maximum likelihood (ML) solution because the prior is flat. If we have more specific information about the solutions we can include them to the problem by choosing an according prior. For testable information it has been proven that the signal power prior is the only valid prior from probability theory's point of view. It consists of a Gaussian with exponent α (regularization parameter) where the solution $\mathbf{x}$ is weighted by a matrix B. Also in the MRI community a bit more specific version of the signal power prior is used: the Tikhonov regularization approach ($B = 1$). The prior does not depend on the model or the covariance, so we can omit them. It writes:

$$\begin{aligned} p(\mathbf{y}|\mathbf{x},\mathcal{A},\mathcal{C}) &= e^{-\frac{1}{2}\left(\mathbf{y}-\mathcal{A}\mathbf{x}\right)^\dagger C^{-1}\left(\mathbf{y}-\mathcal{A}\mathbf{x}\right)} \\ p(\mathbf{x}) &= e^{-\alpha\mathbf{x}^\dagger B\mathbf{x}} \end{aligned} \tag{3.6}$$

Finally this is the posterior PDF that can be used to compute the MAP-solution. For the signal power prior as well as for the flat prior this is possible analytically, whereas in a more

general case a number of samples can be generated by a Markov Chain Monte Carlo run. Interestingly, the Bayesian interpretation has been proven to be the only correct interpretation of tackling a problem. Any other approach is either wrong or equivalent. That leads us to an interesting connection to least squares optimization which is just an equivalent formulation of the Bayesian picture.

3.3 Inverse problems as optimization problems

3.3.1 Weighted complex least squares optimization

In section 3.1.1 about the pseudo inverse only problems with ideal noiseless measurements have been dealt with. For any real physical application the data is influenced by a certain kind of additive noise. Following [17], our problem writes as

$$\mathbf{y} = \mathcal{A}\mathbf{x} + \mathbf{n}$$

The noise $\mathbf{n}$ be of multivariate Gaussian type with known covariance matrix $C = \mathrm{cov}(\mathbf{n})$. This information about the noise distribution can be used to find a solution of the linear system. In the noiseless case the solution was computed straight-forwardly with the help of the pseudo-inverse and the SVD as long as the problem had full rank. For problems that do not have full rank anymore a common approach is to solve them in the least squares sense. This is achieved by minimizing the residual function $\Phi(\mathbf{x})$ which is the weighted sum of squares of the components of $\mathbf{n}$. From equation (3.1) follows that $\mathbf{n} = \mathbf{y} - \mathcal{A}\mathbf{x}$ which are weighted with the inverse of the covariance matrix C. Then the residual function $\Phi(\mathbf{x})$ reads

$$\Phi(\mathbf{x}) = \big(\mathbf{y} - \mathcal{A}\mathbf{x}\big)^\dagger C^{-1} \big(\mathbf{y} - \mathcal{A}\mathbf{x}\big) \tag{3.7}$$

and the solution of the linear system in weighted least squares sense is

$$\hat{\mathbf{x}} = \min_{\mathbf{x}} \Big\{ \Phi(\mathbf{x}) \Big\} = \min_{\mathbf{x}} \Big\{ \big(\mathbf{y} - \mathcal{A}\mathbf{x}\big)^\dagger C^{-1} \big(\mathbf{y} - \mathcal{A}\mathbf{x}\big) \Big\}$$

This is exactly the same problem formulation like in the Bayesian picture: Φ is the exponent of the likelihood function in equation 3.5 up to the factor $1/2$ - which has no importance for the optimization. Minimizing Φ is the same as maximizing the likelihood PDF. The analytic solution of above's equation (which is the ML-solution) can be computed by requiring the gradient vector of Φ with respect to $\mathbf{x}$ to be the zero vector.

$$\nabla_{\mathbf{x}} \Phi(\mathbf{x}) = -\big(\mathbf{y}^\dagger C^{-1} - \mathbf{x}^\dagger \mathcal{A}^\dagger C^{-1}\big)\mathcal{A} \stackrel{!}{=} \mathbf{0}$$

Because the the first derivative is the complex conjugate of the second, either one will be zero if the other one is. C is a covariance matrix and therefore $C^{-1} = (C^{-1})^\dagger$ holds. If $\mathcal{A}^\dagger\mathcal{A}$ is positive definite such that Φ has positive sign for all nonzero $\mathbf{x}$, the solution $\hat{\mathbf{x}}$ minimizes f and is not a saddle point. Using the complex conjugate of the gradient we get the **backprojected problem**:

$$\boxed{\mathcal{A}^\dagger C^{-1} \mathcal{A}\mathbf{x} = \mathcal{A}^\dagger C^{-1} \mathbf{y}} \tag{3.8}$$

So far no constraints whether $\mathcal{A}$ has full rank or not have been made. In the following these two cases will be discussed.

Full rank

To derive an explicit solution for above's equation $\mathcal{A}^\dagger C^{-1}\mathcal{A}$ needs to be invertable (has full rank) such that $(\mathcal{A}^\dagger C^{-1}\mathcal{A})^{-1}$ exists. This is the case when the columns of $C^{-\frac{1}{2}}\mathcal{A}$ are linear independent which means $C^{-\frac{1}{2}}\mathcal{A}$ has full column rank. Then $\mathcal{A}^\dagger C^{-\frac{1}{2}}$ is also invertable and has full rank. The following conditions hold:

1. $(\mathcal{A}^\dagger C^{-1}\mathcal{A})^{-1} = (\mathcal{A}^\dagger C^{-1}\mathcal{A})^{+}$
2. $(\mathcal{A}^\dagger C^{-\frac{1}{2}})^{-1} = (\mathcal{A}^\dagger C^{-\frac{1}{2}})^{+}$
3. $(C^{-\frac{1}{2}}\mathcal{A})^{-1} = (C^{-\frac{1}{2}}\mathcal{A})^{+}$
4. $C^{-\frac{1}{2}}\mathcal{A}(C^{-\frac{1}{2}}\mathcal{A})^{+} = \mathbf{1}$

The explicit solution $\hat{\mathbf{x}}$ of the weighted complex least squares problem is then uniquely given by multiplying equation (3.8) with $(\mathcal{A}^\dagger C^{-1}\mathcal{A})^{-1}$ from the left. It can be shown that this is equivalent with multiplying with $(C^{-\frac{1}{2}}\mathcal{A})^{+}$.

$$
\begin{aligned}
\hat{\mathbf{x}} &= (\mathcal{A}^\dagger C^{-1}\mathcal{A})^{-1}\mathcal{A}^\dagger C^{-1}\mathbf{y} \\
&= (\mathcal{A}^\dagger C^{-\frac{1}{2}}C^{-\frac{1}{2}}\mathcal{A})^{-1}\mathcal{A}^\dagger C^{-\frac{1}{2}}C^{-\frac{1}{2}}\mathbf{y} && \textit{Cholesky} \\
&= (C^{-\frac{1}{2}}\mathcal{A})^{-1}(\mathcal{A}^\dagger C^{-\frac{1}{2}})^{-1}\mathcal{A}^\dagger C^{-\frac{1}{2}}C^{-\frac{1}{2}}\mathbf{y} && (AB)^{-1} = B^{-1}A^{-1} \\
&= (C^{-\frac{1}{2}}\mathcal{A})^{+}\underbrace{(\mathcal{A}^\dagger C^{-\frac{1}{2}})^{+}\mathcal{A}^\dagger C^{-\frac{1}{2}}}_{\mathbf{1}\text{ because of 4.}}C^{-\frac{1}{2}}\mathbf{y} && (1) \\
&= (C^{-\frac{1}{2}}\mathcal{A})^{+}C^{-\frac{1}{2}}\mathbf{y}
\end{aligned}
$$

with the pseudo-inverse defined as:

$$
(C^{-\frac{1}{2}}\mathcal{A})^{+} = (\mathcal{A}^\dagger C^{-1}\mathcal{A})^{-1}\mathcal{A}^\dagger C^{-\frac{1}{2}} \tag{3.9}
$$

The pseudo-inverse of $C^{-\frac{1}{2}}\mathcal{A}$ can be computed efficiently with the help of SVD if the matrix size is not too big. A remarkable property of the weighted least-squares solution $\hat{\mathbf{x}}$ is that it does not only minimize the sum-of-squares functional $\Phi(\mathbf{x})$ but it also solves the original linear system (3.1). Since $(\mathcal{A}^\dagger C^{-1})^{-1}$ exists, the multiplication of equation (3.8) from the left with $(\mathcal{A}^\dagger C^{-1})^{-1}$ proves that statement.

Deficient rank

In the case that the matrix $C^{-\frac{1}{2}}\mathcal{A}$ has linear dependent columns due to redundant information, $C^{-\frac{1}{2}}\mathcal{A}(C^{-\frac{1}{2}}\mathcal{A})^{+}$ is not the identity matrix anymore. The solutions $\mathbf{x}$ of the linear system - if any exist - are not unique. The minimizer $\hat{\mathbf{x}}$ of the weighted least squares functional can still be calculated using the pseudoinverse like in equation (3.9) but it does not solve the original linear system like in the full-rank case. Though it minimizes $\Phi(\mathbf{x})$ and therefore gives a lower bound for all solutions $\mathbf{x}$ of the linear system. In the following a method will be discussed that enables us to convert a rank deficient problem into a full-rank one such that the convenience of using the pseudo-inverse for the solution can still be enjoyed. Of course, that requires to bring further assumptions about the shape of the solution into play.

3.3.2 Tikhonov regularization

A way of transforming a rank-deficient problem into a full-rank problem is to add a regularization term. One of the most common regularizations is the Tikhonov regularization which puts an additional constraint in the form of a matrix T on the solution. This can also be seen as a confinement of the space of solutions to a certain subspace. When $T = \mathbf{1}$ the global smallness of the solution will be emphasized. Another possible way of choosing T is to interpret it as a covariance matrix of the parameters $\mathbf{x}$ and using the correlation between the parameters as prior information if they are known. The regularization parameter α weights the Tikhonov-constraint compared to the data fidelity term for the weighted least squares case. A way of choosing the optimal α in the sense of Bayesian statistics is presented in one of the following chapters. The altered minimization problem reads

$$\hat{\mathbf{x}}_{Tik} = \min_{\mathbf{x}} \Big\{ \big(\mathbf{y} - \mathcal{A}\mathbf{x}\big)^\dagger C^{-1} \big(\mathbf{y} - \mathcal{A}\mathbf{x}\big) + \alpha \|T\mathbf{x}\|_2^2 \Big\}$$

for $\alpha > 0$ and $\alpha \in \mathbf{R}$. The residual function for a general Tikhonov regularization with $B := T^\dagger T$ is

$$\Phi(\mathbf{x}) = \big(\mathbf{y} - \mathcal{A}\mathbf{x}\big)^\dagger C^{-1} \big(\mathbf{y} - \mathcal{A}\mathbf{x}\big) + \alpha \mathbf{x}^\dagger B \mathbf{x} \tag{3.10}$$

This cost function is - up to the factor 1/2 - again exactly the same as the exponent for the posterior PDF for the signal power prior in equation 3.6. Writing this cost function similarly to equation (3.8) yields the full rank **regularized backprojected problem**.

$$\boxed{\big(\mathcal{A}^\dagger C^{-1} \mathcal{A} + \alpha B\big)\mathbf{x} = \mathcal{A}^\dagger C^{-1} \mathbf{y}} \tag{3.11}$$

Rearranging the matrices in the weighted least squares problem above yields an equation where the solution can be extracted in analogy to the unregularized problem.

$$\hat{\mathbf{x}}_{Tik} = \min_{\mathbf{x}} \left\{ \left(\begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} - \begin{pmatrix} \mathcal{A} \\ \sqrt{\alpha} T \end{pmatrix} \mathbf{x} \right)^\dagger \begin{pmatrix} C^{-1} \\ \mathbf{0} \end{pmatrix} \left(\begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} \right) - \begin{pmatrix} \mathcal{A} \\ \sqrt{\alpha} T \end{pmatrix} \mathbf{x} \right\}$$

Therefore the explicit solution of the weighted least squares problem with Tikhonov regularization is:

$$\hat{\mathbf{x}}_{Tik} = \big(\mathcal{A}^\dagger C^{-1} \mathcal{A} + \alpha B\big)^{-1} \mathcal{A}^\dagger C^{-1} \mathbf{y} \tag{3.12}$$

The inverse matrix in above's equation can be computed either with the SVD or directly because it has full rank. The solution is equivalent to the MAP solution of equation 3.6.

3.3.3 Conjugate gradient

Computing the pseudo-inverse matrix in equation (3.12) gets exhausting for high-dimensional problems because direct inversion as well as SVD requires $O(n^3)$ operations for a $n \times n$ matrix which can soon require too much time or memory. Since equation (3.12) is still a convex optimization problem, any iterative algorithm for solving this class of problems can be applied instead. We will concentrate on the conjugated gradient method here because it is used in the numerical SENSE-implementation that was used in the numerical studies. The conjugate gradient algorithm minimizes cost functions of the type $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\dagger A \mathbf{x} - \mathbf{b}^\dagger \mathbf{x} + c$

(as in [6]). Comparing that with the residual function Φ in equation (3.10) for the Tikhonov regularized least squares case, we identify

$$A = \mathcal{A}^\dagger C^{-1} \mathcal{A} + \alpha B$$
$$\mathbf{b} = \mathcal{A}^\dagger C^{-1} \mathbf{y}$$
$$c = \frac{1}{2} \mathbf{y}^\dagger C^{-1} \mathbf{y}$$

Minimizing Φ is equivalent to solving the backprojected problem (3.11), which reads $A\mathbf{x} = \mathbf{b}$ with above's definitions. The conjugate gradient method is related to the steepest descent method where the minimum of f is found by searching along the direction of the steepest descent $\mathbf{r}_k = -\nabla f = \mathbf{b} - A\mathbf{x}_k$. However, this method can have a very slow convergence for ill-conditioned problems. If the condition number of A is high, there is a wide range between small and big eigenvalues. This asymmetry also shows in the plot of the contour lines of f which have a strong non-spherical shape such that the minimum is approached very slowly in zig-zag lines. The idea of conjugate gradient is now that all consecutive search directions $\mathbf{d}_0, \mathbf{d}_1, ..., \mathbf{d}_{N-1}, \mathbf{d}_N$ have to be orthogonal and therefore are not necessarily equivalent to the negative gradient of f. The solution is formed in a Krylov space formed by the vectors $\mathbf{d}_k$ and the coefficients λ_k. The residual at step k is the deviation of the current solution from the $\mathbf{b}$-vector. A new basis vector $\mathbf{d}_{k+1}$ of the Krylov space is found by taking a combination of the previous basis vector and the residual where the factor β_k ensures orthogonality. Convergence can be detected for example by setting a lower bound for $|\mathbf{r}_k|$.

Algorithm 1: Conjugate gradient algorithm

Choose $\mathbf{x}_0$ as initial vector.
Choose $\mathbf{r}_0 := \mathbf{b} - A\mathbf{x}_0$ as initial residual.
Set $\mathbf{d}_0 := \mathbf{r}_0$ as initial search direction.
$k \leftarrow 1$
for $k = 1$ *to* N *do* **do**
  $\mathbf{a}_k = A\mathbf{d}_n$
  $\lambda_k \leftarrow \frac{\mathbf{r}_k^\dagger \mathbf{r}_k}{\mathbf{d}_k^\dagger \mathbf{a}_k}$
  $\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k + \lambda_k \mathbf{d}_k$
  $\mathbf{r}_{k+1} \leftarrow \mathbf{r}_k - \lambda_k \mathbf{x}_k$
  $\beta_k \leftarrow \frac{\mathbf{r}_{k+1}^\dagger \mathbf{r}_{k+1}}{\mathbf{r}_k^\dagger \mathbf{r}_k}$
  $\mathbf{d}_{k+1} \leftarrow \mathbf{r}_{k+1} + \beta_{k+1} \mathbf{d}_k$
  if *Converged* **then**
    Stop
  end
  $k \leftarrow k + 1$
end

3.4 Markov chain Monte Carlo (MCMC)

MCMC is a technique to sample from a multidimensional PDF by generating consecutive configurations from a Markov chain whose stationary distribution corresponds to the desired distribution. The sampled configurations will be highly correlated and especially dependent on the initial state so convergence needs to be monitored carefully to obtain solid results.

The sampling itself can be implemented by different algorithms. If the PDF that is sampled from can be split into all its conditional distributions, a Gibbs-sampler is applicable (compare [13]). That has the additional advantage that it saves computational costs, because it can be paralleled. Unfortunately in the case of MRI Gibbs-sampling is not possible because the model contains a Fourier transform which correlates each pixel with all the others. Therefore another sampling algorithm is used which is presented in the following section.

3.4.1 The Metropolis-Hastings algorithm

The Metropolis-Hastings method has been used in statistical physics to compute the equilibrium distribution of a 2D ferromagnetic spin lattice according to the Boltzmann distribution. Since an MR image can also be seen as a 2D object with an arbitrary number of possible states per pixel (in contrast to 2 spin states in the lattice), there is already an intuitive analogy between the two topics. In the first the energy at a certain temperature according to the interactions of the spins in the lattice is used as cost function. In MRI this place is taken by the cost function that contains likelihood and prior function.

The algorithm starts by evolving from an initial state $\mathbf{x}_0$ and explores the state space by proposing new states according to a proposal probability $q(\mathbf{x}_a \rightarrow \mathbf{x}_n)$. The proposal PDF has to be chosen in such a way that the Markov chain is ergodic (that is irreducible and aperiodic), because otherwise certain parts of the state space might not be reached and the chain does not converge to p. Each new state $\mathbf{x}_n$ will be accepted with the probability

$$p_{acc}(\mathbf{x}_n) = \min\left(1, \frac{p(\mathbf{x_n})\ q(\mathbf{x}_n \rightarrow \mathbf{x}_a)}{p(\mathbf{x_a})\ q(\mathbf{x}_a \rightarrow \mathbf{x}_n)}\right) \tag{3.13}$$

which satisfies detailed balance. That ensures that also a certain ratio of new states with lower probability than their predecessors are going to be accepted to give the chain the chance to escape from local maxima. Choosing a Gaussian PDF (with a certain standard deviation that defines the efficiency of the sampler) for the proposal PDF makes sure that the chain is also ergodic. Therefore - after a sufficient amount of steps - the chain will converge to the desired

stationary distribution p.

Algorithm 2: Metropolis-Hastings algorithm

Define proposal distribution $q(\mathbf{x}_a \to \mathbf{x}_n)$.
Define target distribution $p(\mathbf{x})$.
Set maximal number of states K.
Choose $\mathbf{x}_0$ at random as initial state.
$k \leftarrow 0$
while $k \geq K$ **do**
  Set old state $\mathbf{x}_a \leftarrow \mathbf{x}_k$
  Draw new proposal state $\mathbf{x}_n$ from $q(\mathbf{x}_a \to \mathbf{x}_n)$
  Compute acceptance probability $p_{acc}(\mathbf{x}_n) = \min\left(1, \frac{p(\mathbf{x_n})\ q(\mathbf{x}_n \to \mathbf{x}_a)}{p(\mathbf{x_a})\ q(\mathbf{x}_a \to \mathbf{x}_n)}\right)$
  if $p_{acc}(\mathbf{x}_n) \geq 1$ **then**
    Accept proposal state $\mathbf{x}_{k+1} = \mathbf{x}_n$
  else
    Draw $r \propto$ Uniform(0,1)
    if $p_{acc}(\mathbf{x}_n) \geq r$ **then**
      Accept proposal state $\mathbf{x}_{k+1} = \mathbf{x}_n$
    else
      Keep old state $\mathbf{x}_{k+1} = \mathbf{x}_a$
    end
  end
  $k \leftarrow k + 1$
end

3.4.2 Error estimation and convergence

The initial state of a MCMC run can be chosen for example at random or by evaluating a simple approximation of the model. Then the sampler is started and the chain will move towards the stationary distribution, if the chain is ergodic and satisfies detailed balance. The first samples will be highly dependent on the initial state. Therefore it is important to discard those correlated since they carry less information about the target distribution than later ones. Usually the first $20 - 50\%$ of a MCMC run are ignored. After the chain has overcome this Burn-in period, statistical inference makes sense and the generated samples can be evaluated. For a detailed discussion see [23].

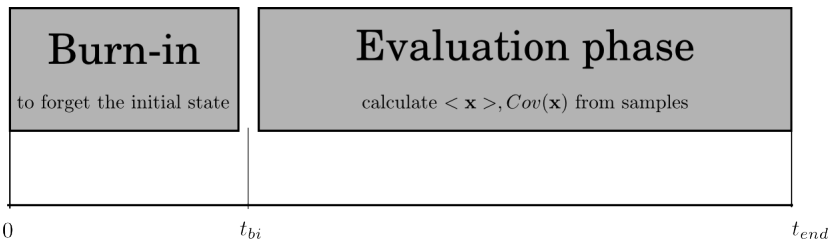


Figure 12: Schematic MCMC-run.

However, waiting for the chain to overcome the Burn-in is not sufficient for convergence. Depending on the proposal distribution and the number of iterations the chain might still be far off of the stationary distribution. Any measurements taken in this situation would be useless and lead to wrong results. That is why convergence and autocorrelation of the chain have to be monitored thoroughly. In the following two approaches are presented that will help to check the correlation of the samples and how far the chain is away from convergence.

Autocorrelation function
This part is mostly inspired by the lecture notes [6]. The autocorrelation function can be used to check if the chain is sufficiently burnt in and it is used to correct the error for an observable. It basically describes the correlation of N realizations $X_1, X_2, ..., X_N$ of a (one-dimensional) stochastic process X_j at time j and $j+t$ separated by a time lag t. The autocorrelation function $\rho(t)$:$\mathrm{R}^+ \to [-1, 1]$ is defined as the covariance between X_j and X_{j+t}, where $\sigma_j = \sqrt{Var\big(X_j\big)}$:

$$\begin{aligned} \rho(t) :&= \frac{Cov\big(X_j, X_{j+t}\big)}{\sigma_j \sigma_{j+t}} \\ &= \frac{\mathrm{E}\big[\big(X_j - \mathrm{E}\big[X_j\big]\big)\big(X_{j+t} - \mathrm{E}\big[X_{j+t}\big]\big)\big]}{\sigma_j \sigma_{j+t}} \end{aligned} \tag{3.14}$$

Given a sufficiently big number N of realizations of X_j an estimate for the expectation value $\mathrm{E}\big[X_j\big]$ is found in the average value. For convenience a new random variable $Y_j = X_{j+t}$ is introduced. Indicating the t-dependence by explicitly writing it as a function of time the averages are defined as:

$$\overline{X}(t) = \frac{1}{N-t} \sum_{j=1}^{N-t} X_j$$

$$\overline{Y}(t) = \frac{1}{N-t} \sum_{j=1}^{N-t} Y_j = \frac{1}{N-t} \sum_{j=t+1}^{N} X_j$$

The estimators for the standard deviations σ_j and σ_{j+t} become

$$\sigma_j(t) = \sqrt{\sum_{j=1}^{N-t} \big(X_j \ - \overline{X}(t)\big)^2}$$

$$\sigma_{j+t}(t) = \sqrt{\sum_{j=1}^{N-t} \big(Y_j \ - \overline{Y}(t)\big)^2}$$

Replacing means and variances in equation 3.14 with their estimators yields and expression for the estimator for the autocorrelation function:

$$\rho^E(t) := \frac{\sum_{j=1}^{N-t} \big(X_j - \overline{X}(t)\big)\big(Y_j - \overline{Y}(t)\big)}{\sqrt{\left(\sum_{j=1}^{N-t} \big(X_j \ - \overline{X}(t)\big)^2\right)\left(\sum_{j=1}^{N-t} \big(Y_j \ - \overline{Y}(t)\big)^2\right)}}$$

This result can be generalized directly from one-dimensional to M-dimensional random variables $\mathbf{X}_j$ and for this case it looks like

$$\rho^E(t) := \frac{\sum_{j=1}^{N-t} \big(\mathbf{X}_j - \overline{\mathbf{X}}(t)\big)^T\big(\mathbf{Y}_j - \overline{\mathbf{Y}}(t)\big)}{\sqrt{\left(\sum_{j=1}^{N-t} \big(\mathbf{X}_j \ - \overline{\mathbf{X}}(t)\big)^2\right)\left(\sum_{j=1}^{N-t} \big(\mathbf{Y}_j \ - \overline{\mathbf{Y}}(t)\big)^2\right)}} \tag{3.15}$$

Equivalently the averages are defined componentwise:

$$\overline{\mathbf{X}}(t) = \frac{1}{N-t}\sum_{j=1}^{N-t} \mathbf{X}_j$$

$$\overline{\mathbf{Y}}(t) = \frac{1}{N-t}\sum_{j=1}^{N-t} \mathbf{Y}_j = \frac{1}{N-t}\sum_{j=t+1}^{N} \mathbf{X}_j$$

$\rho^E(t)$ is also defined in the interval $[-1, 1]$. Negative values indicate that $\mathbf{X}_j$ and $\mathbf{X}_{j+t}$ are anti-correlated, positive values indicate correlation which is just another way of describing a (in-)direct relationship of two variables. In general, the absolute value of the estimated autocorrelation function (which will be referred to as autocorrelation function for the sake of brevity) is the property that is most interesting since it serves as a measure for the correlation of samples in a time series. For the establishment of solid averages from the Markov chain - and therefore for the extraction of a useful result from our simulation - the generated samples need to be as uncorrelated as possible. Observing the evolution of the autocorrelation function live during the simulation gives an idea when a sufficiently small correlation of the samples has been reached. Computing the autocorrelation live during the simulation has to be as computationally efficient as possible to keep simulation time to a minimum. In 6.1 in the appendix more detailed considerations for efficient evaluation of $\rho^E(t)$ can be found.

A Markov chain has always a certain amount of autocorrelation - even after the chain has forgotten its initial state - because each step depends on the one before. This fact has to be taken into account when computing errors and is quantified by using the integrated autocorrelation time τ_{int}.

$$\mathrm{Cov}\big(\mathbf{X}\big) = \frac{\langle \mathbf{X}\mathbf{X}^\dagger\rangle - \langle\mathbf{X}\rangle\langle\mathbf{X}\rangle^\dagger}{N}\ 2\tau_{int}$$

$$\tau_{int} := \frac{1}{2} + \sum_{t=1}^{N} \rho_E(t)\Big(1 - \frac{t}{N}\Big)$$

The factor $2\tau_{int}$ is equal to one if the samples were uncorrelated. The autocorrelation function can be expanded as a series of decaying exponentials $\exp(-t/\tau_i)$ over simulation time. The

integrated autocorrelation time is approximately τ_1. That is why it has to be taken care that the simulation time t_end is bigger than the integrated autocorrelation time. Otherwise the Markov chain is still very correlated.

Parallel chains and cross-chain variance
A very simple and useful way to check how far a chain is away from convergence is to run multiple parallel chains. This idea is presented in the Handbook of Markov Chain Monte Carlo by Gelman and Shirley [1] but roots back to Fosdick in 1959. The main idea is that chains starting from different initial states are going to converge to the same target distribution. Depending on the starting configuration some chains might converge faster than others.

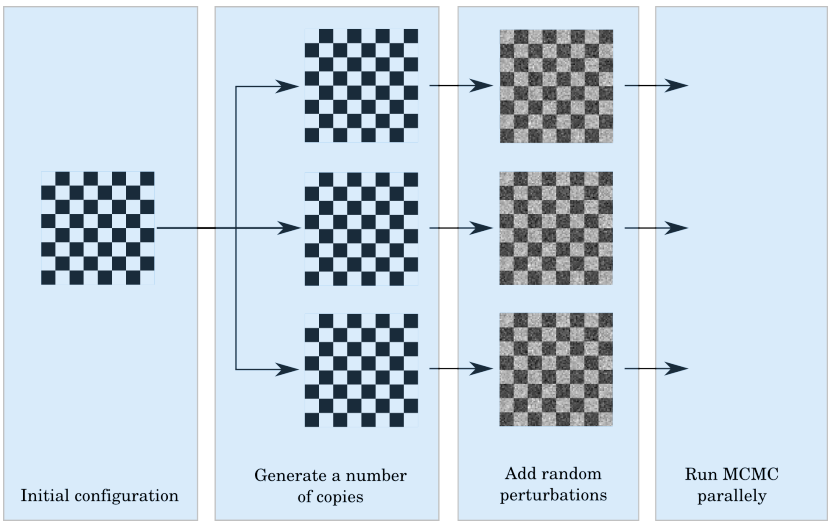


Figure 13: Parallel chains: A single initial guess can be replicated and altered with noise to obtain different starting configurations. Then for each chain the MCMC run can be started.

The variances $\sigma^2_{O^{(i)}}$ of a certain observable at the i^{th} chain (denoted by $O^{(i)}$) can be averaged over all c chains (σ^2_{av}) and compared to the variance of the same observable O computed from all chains mixed together (σ^2_{mix}). The fraction D is equal to 1 if convergence is reached.

$$\sigma^2_{mix} := \langle O^2 \rangle - \langle O \rangle \langle O \rangle$$
$$\sigma^2_{av} := \frac{1}{c} \sum_{i=1}^{c} \sigma^2_{O^{(i)}}$$
$$D := \frac{\sigma^2_{av}}{\sigma^2_{mix}}$$

One single chain can still be very close to the initial state and therefore it might have a considerably "good" (small) variance which can be misleading. Taking multiple chains is useful to provide graphical checks by monitoring the movement of many chains started at different states. That is necessary to prevent the case of having multiple chains with the same variance but at totally different regions in search space (D would also be approximately 1!). It can also help to find bugs in the acceptance rate computation. Running parallel chains has an additional advantage: they generate more samples at once so the simulation time after overcoming the Burn-in period is reduced. More detailed information on Burn-in selection and on an alternative stopping criterion for the special case of image denoising based on MCMC

is found in the PHD-thesis of Cecile Louchet [15].

Optimal step size for the proposal distribution

The convergence rate of a Markov chain depends on the proposal distribution $q(\mathbf{x}_a \rightarrow \mathbf{x}_n)$. If its steps are too big, most of the new states will be rejected, whereas if they are too small the chain will only be able to explore a small part of the search space. Usually it is a common practice to let the chain make rather big moves in the Burn-in phase such that it forgets about the initial state quickly and is able to explore large parts of the space of possible solutions with low acceptance rate (compare [9]). A good strategy to guess the step size is to reconstruct the dataset with conjugate gradient (or any another optimization algorithm like primal dual etc.) which is not costly. The pixel values of this reconstruction can be used to make a guess of the expected order of magnitude. For the step size during Burn-in one can use 10 or 20 % of this characteristic value. After Burn-in the step size needs to be even smaller such that the acceptance rate is around 23 %. This optimal scaling of the proposal step size is derived in *Rosenthal and Roberts* ([8]).

4 Methods

In this section methods that try to answer the questions raised in the introduction 1 are presented, in particular for question (2) by finding an optimal regularization parameter and for question (3) by presenting a relative quality measure for reconstructed images by obtaining the covariance matrix of the posterior PDF from two different methods.

4.1 Optimal choice of regularization parameter

How to choose the parameter for regularization is a question that is often avoided answering. From the Bayesian point of view it can be tackled in a very general way that only takes available information into account. That means it only focuses on the structure of the problem itself. In particular it does not even require any further assumptions like human perceptual models or pathological classifications. Given (a) the theoretical model of MRI (e.g. SENSE with coil sensitivities), (b) the measured k-space data, (c) the k-space noise distribution based on the scanner hardware and (d) the regularization method (prior information), the optimal reconstruction parameter can be deduced.

In particular this is achieved by finding an expression for the probability distribution function (PDF) for the regularization parameter α. Via marginalization over all possible images $\mathbf{x}$ and with Bayes theorem the general PDF for α becomes

$$\begin{aligned} p(\alpha|\mathbf{y}, C, \mathcal{I}) &= \int p(\alpha|\mathbf{x}, \mathbf{y}, C, \mathcal{I}) p(\alpha) d^N x \\ &= \frac{1}{Z} \int \underbrace{p(\mathbf{y}|\alpha, \mathbf{x}, C, \mathcal{I})}_{Likelihood} \underbrace{p(\mathbf{x}|\alpha, C, \mathcal{I})}_{Prior} p(\alpha) d^N x \end{aligned}$$

In the above expression two of the three PDFs in the integral are already defined. The Likelihood does not depend on the regularization parameter and hence gets

$$p(\mathbf{y}|\alpha, \mathbf{x}, C, \mathcal{I}) = p(\mathbf{y}|\mathbf{x}, C, \mathcal{I}) = \frac{1}{Z'} e^{-\frac{1}{2}\left(\mathbf{y}-\mathcal{A}\mathbf{x}\right)^\dagger C^{-1}\left(\mathbf{y}-\mathcal{A}\mathbf{x}\right)}$$

The prior is independent of the covariance and the model and yields

$$p(\mathbf{x}|\alpha, C, \mathcal{I}) = p(\mathbf{x}|\alpha) = \frac{1}{Z''(\alpha)} e^{-\alpha R(\mathbf{x})}$$

In the following a quadratic regularization term (Tikhonov, see (4.9)) is assumed which has the form

$$R(\mathbf{x}) = \mathbf{x}^\dagger B \mathbf{x} \tag{4.1}$$

For this quadratic prior the normalization can be computed analytically as

$$\begin{aligned} Z''(\alpha) &= \int e^{-\alpha R(\mathbf{x})} d^N x = \int e^{-\frac{1}{2}\mathbf{x}^\dagger 2\alpha B \mathbf{x}} d^N x \\ &= \sqrt{\frac{(2\pi)^N}{det(2\alpha B)}} = \sqrt{\frac{(2\pi)^N}{(2\alpha)^N det(B)}} \\ &= \left(\frac{\pi}{\alpha}\right)^{\frac{N}{2}} det(B)^{-\frac{1}{2}} \end{aligned}$$

The third - and still unknown - PDF is the prior for the hyperparameter α. Since the latter is a scale parameter - its value depends on the chosen units - and there is no additional information about it, the appropriate prior is Jeffrey's scale prior

$$p(\alpha) = \frac{1}{\alpha} \tag{4.2}$$

Now that all three parts are known the integral can be assembled.

$$\begin{aligned} p(\alpha|\mathbf{y}, C, \mathcal{I}) &= \frac{1}{Z}\int \frac{1}{Z'} e^{-\frac{1}{2}\left(\mathbf{y}-\mathcal{A}\mathbf{x}\right)^\dagger C^{-1}\left(\mathbf{y}-\mathcal{A}\mathbf{x}\right)} \left(\frac{\pi}{\alpha}\right)^{-\frac{N}{2}} det(B)^{\frac{1}{2}} e^{-\alpha R(\mathbf{x})} \frac{1}{\alpha} d^N x \\ &= \frac{1}{\tilde{Z}} \left(\frac{\pi}{\alpha}\right)^{-\frac{N}{2}} det(B)^{\frac{1}{2}} \int e^{-\frac{1}{2}\left(\mathbf{y}-\mathcal{A}\mathbf{x}\right)^\dagger C^{-1}\left(\mathbf{y}-\mathcal{A}\mathbf{x}\right) - \alpha R(\mathbf{x})} \frac{1}{\alpha} d^N x \\ &= \frac{1}{\tilde{Z}} \alpha^{\frac{N-2}{2}} \int e^{-\frac{1}{2}\Phi(\mathbf{x},\alpha)} d^N x \end{aligned}$$

For the quadratic image prior in equation (4.1) the integral above is Gaussian and can be computed analytically (compare section 3.3.2 on weighted complex LSQ with Tikhonov regularization).

$$\begin{aligned} \Phi(\mathbf{x}, \alpha) &:= \left(\mathbf{y} - \mathcal{A}\mathbf{x}\right)^\dagger C^{-1} \left(\mathbf{y} - \mathcal{A}\mathbf{x}\right) + 2\alpha R(\mathbf{x}) \\ &= \mathbf{y}^\dagger C^{-1}\mathbf{y} - 2\mathbf{x}^\dagger \mathcal{A}^\dagger C^{-1}\mathbf{y} + \mathbf{x}^\dagger \underbrace{\left(\mathcal{A}^\dagger C^{-1}\mathcal{A} + 2\alpha B\right)}_{=:H(\alpha)} \mathbf{x} \end{aligned} \tag{4.3}$$

$$\text{with} \qquad \boxed{H(\alpha) := \mathcal{A}^\dagger C^{-1}\mathcal{A} + 2\alpha B}$$

To compute the integral in general a saddle point approximation is applied to the exponent. This can be done for general regularizations $R(\mathbf{x})$, for the quadratic (Tikhonov) it is exact when breaking the expansion after second order terms. The exponent $\Phi(\mathbf{x}, \alpha)$ is approximated by a Taylor series around its global minimum $\mathbf{x}^*$. Since H has full rank it can be computed by inversion.

$$\begin{aligned} \nabla_{\mathbf{x}} \Phi(\mathbf{x}, \alpha) &= 2\left(H\mathbf{x} - \mathcal{A}^\dagger C^{-1}\mathbf{y}\right) = 0 \\ \mathbf{x}^* &= H^{-1}\mathcal{A}^\dagger C^{-1}\mathbf{y} \end{aligned}$$

Plugging the global minimum into the second-order Taylor approximation for Φ gives the following expression

$$\begin{aligned} \Phi(\mathbf{x}, \alpha) = & \underbrace{\Phi(\mathbf{x}^*, \alpha)}_{\mathbf{y}^\dagger C^{-1}\mathbf{y} - 2\mathbf{x}^{*\dagger}\mathcal{A}^\dagger C^{-1}\mathbf{y} + \mathbf{x}^{*\dagger} H \mathbf{x}^*} + (\mathbf{x} - \mathbf{x}^*)^\dagger \underbrace{\nabla_{\mathbf{x}} \Phi(\mathbf{x}, \alpha)\big|_{\mathbf{x}=\mathbf{x}^*}}_{2\left(H\mathbf{x}^* - \mathcal{A}^\dagger C^{-1}\mathbf{y}\right)} + \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^\dagger \underbrace{\nabla_{\mathbf{x}}\nabla_{\mathbf{x}} \Phi(\mathbf{x}, \alpha)\big|_{\mathbf{x}=\mathbf{x}^*}}_{=2H} (\mathbf{x} - \mathbf{x}^*) \\ = & \mathbf{y}^\dagger C^{-1}\mathbf{y} - \mathbf{x}^{*\dagger} H \mathbf{x}^* - 2\mathbf{x}^\dagger \mathcal{A}^\dagger C^{-1}\mathbf{y} + 2\mathbf{x}^\dagger H \mathbf{x}^* + (\mathbf{x} - \mathbf{x}^*)^\dagger H (\mathbf{x} - \mathbf{x}^*) \\ = & \mathbf{y}^\dagger C^{-1}\mathbf{y} - \mathbf{y}^\dagger C^{-1}\mathcal{A}\underbrace{H^{-1\dagger} H}_{\mathbf{1}} H^{-1}\mathcal{A}^\dagger C^{-1}\mathbf{y} - 2\mathbf{x}^\dagger \mathcal{A}^\dagger C^{-1}\mathbf{y} + 2\mathbf{x}^\dagger \underbrace{H H^{-1}}_{\mathbf{1}} \mathcal{A}^\dagger C^{-1}\mathbf{y} + \\ & + (\mathbf{x} - \mathbf{x}^*)^\dagger H (\mathbf{x} - \mathbf{x}^*) \\ = & \mathbf{y}^\dagger C^{-1}\mathbf{y} - \mathbf{y}^\dagger C^{-1}\mathcal{A} H^{-1}\mathcal{A}^\dagger C^{-1}\mathbf{y} + (\mathbf{x} - \mathbf{x}^*)^\dagger H (\mathbf{x} - \mathbf{x}^*) \end{aligned}$$

Now this (exact!!) approximation can be plugged into the integral again, which reads

$$\begin{aligned} p(\alpha|\mathbf{y},C,\mathcal{I}) &= \frac{1}{\bar{Z}}\alpha^{\frac{N-2}{2}}\int e^{-\frac{1}{2}\Phi(\mathbf{x},\alpha)}d^N x \\ &= \frac{1}{\bar{Z}}\alpha^{\frac{N-2}{2}}\int e^{-\frac{1}{2}\mathbf{y}^\dagger C^{-1}\mathbf{y}}e^{\frac{1}{2}\mathbf{y}^\dagger C^{-1}\mathcal{A}H^{-1}\mathcal{A}^\dagger C^{-1}\mathbf{y}}e^{-\frac{1}{2}(\mathbf{x}-\mathbf{x}^*)^\dagger H(\mathbf{x}-\mathbf{x}^*)}d^N x \\ &= \frac{1}{\bar{Z}}\alpha^{\frac{N-2}{2}}e^{-\frac{1}{2}\mathbf{y}^\dagger C^{-1}\mathbf{y}}e^{\frac{1}{2}\mathbf{y}^\dagger C^{-1}\mathcal{A}H^{-1}\mathcal{A}^\dagger C^{-1}\mathbf{y}}\underbrace{\int e^{-\frac{1}{2}(\mathbf{x}-\mathbf{x}^*)^\dagger H(\mathbf{x}-\mathbf{x}^*)}d^N x}_{\sqrt{\frac{(2\pi)^N}{det(H)}}} \end{aligned}$$

where $\hat{Z}$ is a normalization constant with all the α-independent terms. The PDF for the regularization parameter α then is:

$$\boxed{p(\alpha|\mathbf{y},C,\mathcal{I}) = \frac{1}{\hat{Z}}\alpha^{\frac{N-2}{2}}e^{\frac{1}{2}\mathbf{y}^\dagger C^{-1}\mathcal{A}H^{-1}\mathcal{A}^\dagger C^{-1}\mathbf{y}}det(H)^{-\frac{1}{2}}} \tag{4.4}$$

To obtain the optimal regularization parameter the MAP-solution of above's PDF is be computed. Therefore the PDF $p(\alpha|\mathbf{y},C,\mathcal{I})$ is reformulated as a log-PDF $L(\alpha)$ which looks like

$$L(\alpha) := ln(p(\alpha|\mathbf{y},C,\mathcal{I})) = C_0 + \frac{N-2}{2}ln(\alpha) + \frac{1}{2}\mathbf{y}^\dagger C^{-1}\mathcal{A}H^{-1}\mathcal{A}^\dagger C^{-1}\mathbf{y} - \frac{1}{2}\underbrace{ln(det(H))}_{tr(ln(H))}$$

$$\begin{aligned} \frac{\partial}{\partial\alpha}L(\alpha) &= \frac{N-2}{2\alpha} + \frac{1}{2}\frac{\partial}{\partial\alpha}\mathbf{y}^\dagger C^{-1}\mathcal{A}H^{-1}\mathcal{A}^\dagger C^{-1}\mathbf{y} - \frac{1}{2}\frac{\partial}{\partial\alpha}tr(ln(H)) = 0 \\ \frac{2-N}{\alpha} &= \frac{\partial}{\partial\alpha}\mathbf{y}^\dagger C^{-1}\mathcal{A}H^{-1}\mathcal{A}^\dagger C^{-1}\mathbf{y} - \frac{\partial}{\partial\alpha}tr(ln(H)) \end{aligned} \tag{4.5}$$

Now the derivatives on the right hand side can be simplified.

$$\begin{aligned} \frac{\partial}{\partial\alpha}\mathbf{y}^\dagger C^{-1}\mathcal{A}H^{-1}\mathcal{A}^\dagger C^{-1}\mathbf{y} &= \mathbf{y}^\dagger C^{-1}\mathcal{A}\frac{\partial}{\partial\alpha}\mathbf{z} \\ \mathbf{z} &:= H^{-1}\mathbf{b} \\ \mathbf{b} &:= \mathcal{A}^\dagger C^{-1}\mathbf{y} \end{aligned}$$

The expression $\frac{\partial}{\partial\alpha}\mathbf{z}$ can be determined from $\mathbf{b} = H\mathbf{z}$. Since $\mathbf{b}$ is independent from α, we can write

$$\begin{aligned} 0 = \frac{\partial}{\partial\alpha}\mathbf{b} &= \Big(\frac{\partial}{\partial\alpha}H\Big)\mathbf{z} + H\frac{\partial}{\partial\alpha}\mathbf{z} \\ \frac{\partial}{\partial\alpha}\mathbf{z} &= -H^{-1}\Big(\frac{\partial}{\partial\alpha}H\Big)\mathbf{z} \\ &= -H^{-1}\underbrace{\Big(\frac{\partial}{\partial\alpha}H\Big)}_{2B}H^{-1}\mathbf{b} \\ &= -2H^{-1}BH^{-1}\mathcal{A}^\dagger C^{-1}\mathbf{y} \end{aligned}$$

Hence the first derivative gets

$$\frac{\partial}{\partial\alpha}\mathbf{y}^\dagger C^{-1}\mathcal{A}H^{-1}\mathcal{A}^\dagger C^{-1}\mathbf{y} = -2\mathbf{y}^\dagger C^{-1}\mathcal{A}H^{-1}BH^{-1}\mathcal{A}^\dagger C^{-1}\mathbf{y} = -2\mathbf{x}^{*\dagger}B\mathbf{x}^*$$

The second derivative we need is

$$\frac{\partial}{\partial\alpha}tr(ln(H)) = tr(\frac{\partial}{\partial\alpha}ln(H)) = 2tr(H^{-1}B)$$

Then the equation to extract the optimal α from is

$$\boxed{\frac{2-N}{\alpha} = -2\mathbf{y}^{\dagger}C^{-1}\mathcal{A}H^{-1}(\alpha)BH^{-1}(\alpha)\mathcal{A}^{\dagger}C^{-1}\mathbf{y} - 2tr\big(H^{-1}(\alpha)B\big)} \tag{4.6}$$

4.2 Methods for obtaining a samples in images space for a statistical quality measure

To assess the quality of an MR image statistically in the Bayesian sense first all the given information needs to be taken into account. This is basically the same information when looking for the best regularization parameter: (a) the theoretical model $\mathcal{A}$ for SENSE (together with coil sensitivities $c(x, y)$), (b) the measured data $\mathbf{y}_0$, (c) the k-space noise distribution based on the scanner hardware in the form of its covariance matrix C and (d) the regularization method (prior information) with the optimal reconstruction parameter α. All the information (apart from the optimal α) can be exported from the MR scanner. To enable data simulation for various cases without the necessity to perform a real scan and to save computational costs whilst the validation of the presented methods the data acquisition process can be simulated. For that one real, fully sampled dataset is taken which is then transformed to image space. There it can be cropped to the desired size and location of the ROI which is then transformed back to k-space. Additionally any type of undersampling can be applied to the data now. The dataset obtained in such a way now is the input for all further methods. At this point the simulated data is equivalent to arbitrary real data because the MRI model is well conditioned for fully sampled data. That is why the simulation procedure does not introduce a bias to the data. In the following figure a schematic representation of the data acquisition is presented.

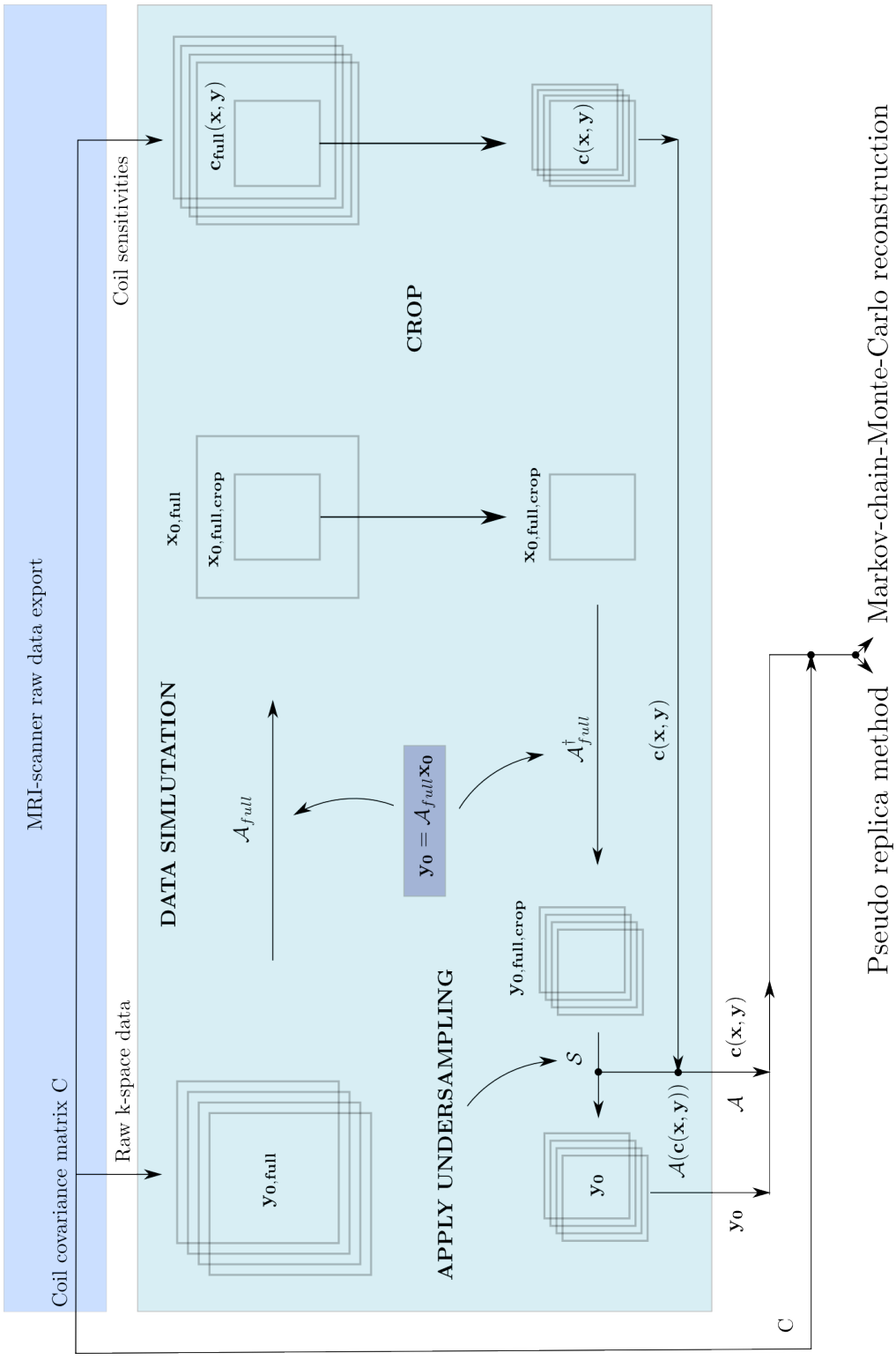


Figure 14: Input data for quality measure methods is either raw MRI-data or a cropped version thereof for flexibility and speedup. It covers: (a) theoretical model $\mathcal{A}$ for SENSE together with coil sensitivities $c(x, y)$, (b) measured data $\mathbf{y}_0$, (c) the k-space noise covariance matrix C and (d) the regularization method reconstruction parameter α.

4.2.1 Pseudo-replica (PR) method

A brute force method to get an error estimation for any image reconstruction algorithm is to take a certain number of independent measurements and apply the algorithm to each of these data sets. The mean and the variance of the outcome will represent a reconstruction with a quality measure which can be used to quantify the quality of the reconstruction for a certain region of pixels. Additionally it can also be used to compare the results of reconstructions from the same data obtained by different algorithms. For linear reconstruction algorithms for MRI-data this method has been presented in [20]. Gaining a sufficient number of independent MRI-datasets is experimentally impractical since it would consume too much time. A work-around for this problem is to take only one measurement of the patient and add artificial noise in order to generate new measurement samples. This works because the noise of the k-space signal is known to be Gaussian noise in real and imaginary part. Measuring the noise distribution of the scanner setup without spatial encoding in a pre-scan and then taking one measurement $\mathbf{y}_0$ of the desired object is therefore enough to reproduce the results of any number of independent measurements. That works as long as the noise has a multivariate Gaussian distribution with unity STD (the distribution is symmetric in all dimensions). From the symmetry between $\mathbf{y}$ and $\mathbf{y}_0$ follows that $\mathcal{N}(\mathbf{y}|\mathbf{y}_0, C) = \mathcal{N}(\mathbf{y}_0|\mathbf{y}, C)$ which allows us to use a single measurement $\mathbf{y}_0$ as the mean of the Gaussian noise distribution and generate PRs $\{\mathbf{y}\}$ according to

$$\mathbf{y} = \mathbf{y}_0 + \mathbf{z}$$

where $\mathbf{z}$ is sampled from the mulivariate normal distribution:

$$p(\mathbf{z}|C) = \frac{1}{Z} e^{-\frac{1}{2}\mathbf{z}^\dagger C^{-1}\mathbf{z}} \tag{4.7}$$

Artificial measurements generated in such a way can be reconstructed by any method. Such a procedure is also called bootstrapping in statistics. From the resulting image samples mean and covariance are computed which can be analyzed to get information about the relative reconstruction quality. The following figure shows a graphical scheme of this concept.

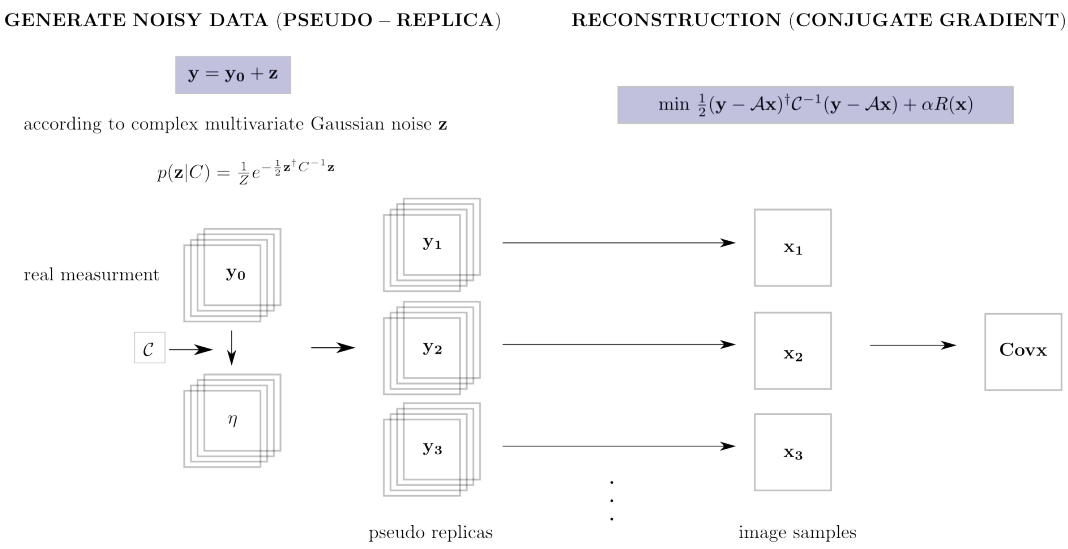


Figure 15: Pseudo replica method: The noise distribution C and a single measurement $\mathbf{y}_0$ are used to generate noisy data samples $\{\mathbf{y}\}$ (pseudo replicas) for the reconstruction. The reconstructed images $\{\mathbf{x}\}$ are used to calculate mean and covariance.

4.2.2 Posterior distribution for general (nonlinear) reconstructions

The target now is to compute the covariance matrix of the posterior PDF given the model for SENSE, the experimental data $\mathbf{y}_0$ acquired in a procedure presented in figure 14, its noise statistics and the regularization method. First of all the expression for the posterior is derived without any assumptions about the noise statistics or the specific kind of the cost function. Let $f : \mathrm{C}^{\mathbf{y}} \to \mathrm{C}^{\mathbf{x}}$ be a linear reconstruction function such that $\mathbf{x} = f(\mathbf{y})$ is the reconstructed image $\mathbf{x}$ from the k-space data $\mathbf{y}$. The density of the reconstructed image $\mathbf{x}$ given the noiseless true data $\mathbf{y}_0$ and any additional information $\mathcal{I}$ (e.g. about the distribution of $\mathbf{y}$) can be calculated by marginalizing over all possible noisy measurements $\mathbf{y}'$.

$$p(\mathbf{x}|\mathbf{y}_0, \mathcal{I}) = \int \underbrace{p(\mathbf{x}|\mathbf{y}', \mathbf{y}_0, \mathcal{I})}_{\delta\left(\mathbf{x} - f(\mathbf{y}')\right)} p(\mathbf{y}'|\mathbf{y}_0, \mathcal{I}) d^N y'$$

The first PDF in the integral is a Dirac-delta centered its unique reconstruction given by f. A variable transform $\mathbf{y}' = f^{-1}(\mathbf{x}')$ in the integral introduces the determinant of the Jacobian matrix $J_{f^{-1}}(\mathbf{x}')$ of the inverse reconstruction function f^{-1}. Then the PDF reads

$$\begin{aligned} p(\mathbf{x}|\mathbf{y}_0, \mathcal{I}) &= \int \delta\big(\mathbf{x} - f(\mathbf{y}')\big) p(\mathbf{y}'|\mathbf{y}_0, \mathcal{I}) d^N y' \\ &= \int \underbrace{\delta\big(\mathbf{x} - f(f^{-1}(\mathbf{x}'))\big)}_{\delta\left(\mathbf{x} - \mathbf{x}'\right)} p(f^{-1}(\mathbf{x}')|\mathbf{y}_0, \mathcal{I}) \underbrace{\left|\frac{df^{-1}(\mathbf{x}')}{d\mathbf{x}'}\right|}_{J_{f^{-1}}(\mathbf{x}')} d^N x' \\ &= p(\mathbf{y}|\mathbf{y}_0, \mathcal{I}) J_{f^{-1}}(\mathbf{x}) \end{aligned} \tag{4.8}$$

This result describes the form of the posterior PDF $p(\mathbf{x}|\mathbf{y}_0, \mathcal{I})$ for images $\mathbf{x}$ given the noiseless data $\mathbf{y}_0$ for any reconstruction method f (e.g a nonlinear one). The PDF $p(\mathbf{y}|\mathbf{y}_0, \mathcal{I})$ describes the noise distribution of the data. In practice the difficulty in this general case is to find the inverse reconstruction function f^{-1} and calculate its Jacobian.

4.2.3 Mean and variance of the posterior distribution for the signal power prior

The PR method can be applied to the signal power regularization approach. This is advantageous because the reconstruction function in this case is linear and the explicit solution for the minimizer of the least squares problem is known. That makes it possible to evaluate mean and covariance of the posterior PDF analytically without actually having to generate pseudo replicas and reconstructing them. Let us take a look at the general expression of the posterior in equation 4.8 now. For a linear reconstruction function the Jacobi determinant $J_{f^{-1}}(\mathbf{x})$ is a constant. If the noise distribution of $\mathbf{y}$ and the Jacobian are known, the posterior can be evaluated directly. However it is more practical to use the general integral expression because it does not require the knowledge of the Jacobian which might be hard to find in practice. The more general expression of equation 4.8 reads

$$p(\mathbf{x}|\mathbf{y}_0, \mathcal{I}) = \int \delta\big(\mathbf{x} - f(\mathbf{y}')\big) p(\mathbf{y}'|\mathbf{y}_0, \mathcal{I}) d^N y'$$

The cost functional is a multivariate Gaussian data fidelity term with an image smallness prior. In the special case of $B = \mathbf{1}$ it is the Tikhonov regularization approach. However, it is

also possible to use other types of priors, for example a discretized first or second derivative matrix for smoothness or curvature. In such a case only the matrix B in the results have to be replaced.

$$\Phi(\mathbf{x}) = \big(\mathbf{y} - \mathcal{A}\mathbf{x}\big)^\dagger C^{-1}\big(\mathbf{y} - \mathcal{A}\mathbf{x}\big) + 2\alpha \mathbf{x}^\dagger B\mathbf{x}$$

The solution that minimizes Φ is given by equation (3.12). Hence the reconstruction function f reads

$$\begin{aligned} f(\mathbf{y}) = \mathbf{x} &= \big(\mathcal{A}^\dagger C^{-1}\mathcal{A} + 2\alpha B\big)^{-1}\mathcal{A}^\dagger C^{-1}\mathbf{y} = H^{-1}\mathcal{A}^\dagger C^{-1}\mathbf{y} \\ H &:= \mathcal{A}^\dagger C^{-1}\mathcal{A} + 2\alpha B \end{aligned}$$

The Likelihood for the data is given by

$$p(\mathbf{y}'|\mathbf{y}_0, \mathcal{I}) = \frac{1}{Z} e^{-\frac{1}{2}(\mathbf{y}'-\mathbf{y}_0)^\dagger C^{-1}(\mathbf{y}'-\mathbf{y}_0)} \tag{4.9}$$

These two specifications are now plugged into in equation (4.9), where the explicit dependence on the covariance matrix C is denoted.

$$p(\mathbf{x}|\mathbf{y}_0, C, \mathcal{I}) = \int \delta\big(\mathbf{x} - f(\mathbf{y}')\big) p(\mathbf{y}'|\mathbf{y}_0, C, \mathcal{I}) d^N y' = \frac{1}{Z}\int \delta\big(\mathbf{x} - f(\mathbf{y}')\big) e^{-\frac{1}{2}(\mathbf{y}'-\mathbf{y}_0)^\dagger C^{-1}(\mathbf{y}'-\mathbf{y}_0)} d^N y'$$

With a coordinate transform $\mathbf{z}' := \mathbf{y}' - \mathbf{y}_0$, using the linearity of f and $f(\mathbf{y}_0) = \mathbf{x}_0$ we obtain

$$p(\mathbf{x}|\mathbf{y}_0, C, \mathcal{I}) = \frac{1}{Z}\int \delta\big(\mathbf{x} - f(\mathbf{z}') - \mathbf{x}_0\big) e^{-\frac{1}{2}\mathbf{z}'^\dagger C^{-1}\mathbf{z}'} d^N z'$$

Now we can calculate the mean as follows

$$\begin{aligned} \langle \mathbf{x} \rangle = \int \mathbf{x}\, p(\mathbf{x}|\mathbf{y}_0, C, \mathcal{I}) d^N x &= \int \mathbf{x} \frac{1}{Z}\int \delta\big(\mathbf{x} - f(\mathbf{z}') - \mathbf{x}_0\big) e^{-\frac{1}{2}\mathbf{z}'^\dagger C^{-1}\mathbf{z}} d^N z' d^N x \\ &= \frac{1}{Z}\int \big(\mathbf{x}_0 + f(\mathbf{z}')\big) e^{-\frac{1}{2}\mathbf{z}'^\dagger C^{-1}\mathbf{z}} d^N z' \\ &= \mathbf{x}_0 + H^{-1}\mathcal{A}^\dagger C^{-1}\frac{1}{Z}\underbrace{\int \mathbf{z}' e^{-\frac{1}{2}\mathbf{z}'^\dagger C^{-1}\mathbf{z}'} d^N z'}_{\mathbf{0}} \\ &= \mathbf{x}_0 \end{aligned}$$

That means that for computing the mean of the PR method no multiple reconstructions are needed. It simply the reconstruction of the experimentally measured dataset $\mathbf{y}_0$. With the definition $\Delta\mathbf{x} := \mathbf{x} - \mathbf{x}_0$ and using the fact that C is symmetric, the covariance can be computed per definition as

$$\begin{aligned} \mathrm{Cov}(\mathbf{x}) = \langle \Delta\mathbf{x}\Delta\mathbf{x}^\dagger \rangle &= \int \big(\mathbf{x} - \mathbf{x}_0\big)\big(\mathbf{x} - \mathbf{x}_0\big)^\dagger \frac{1}{Z}\int \delta\big(\mathbf{x} - f(\mathbf{z}') - \mathbf{x}_0\big) e^{-\frac{1}{2}\mathbf{z}'^\dagger C^{-1}\mathbf{z}'} d^N z' d^N x \\ &= \frac{1}{Z}\int f(\mathbf{z}') f(\mathbf{z}')^\dagger e^{-\frac{1}{2}\mathbf{z}'^\dagger C^{-1}\mathbf{z}'} d^N z' \\ &= H^{-1}\mathcal{A}^\dagger C^{-1}\underbrace{\frac{1}{Z}\int \mathbf{z}'\mathbf{z}'^\dagger e^{-\frac{1}{2}\mathbf{z}'^\dagger C^{-1}\mathbf{z}'} d^N z'}_{C}\big(H^{-1}\mathcal{A}^\dagger C^{-1}\big)^\dagger \\ &= H^{-1}\mathcal{A}^\dagger C^{-1} C \big(H^{-1}\mathcal{A}^\dagger C^{-1}\big)^\dagger \\ &= H^{-1}\mathcal{A}^\dagger C^{-1} C C^{-1}\mathcal{A} H^{-1} \\ &= H^{-1}\underbrace{\mathcal{A}^\dagger C^{-1}\mathcal{A}}_{H - 2\alpha B} H^{-1} \\ &= H^{-1} - H^{-1} B H^{-1} \end{aligned}$$

The last expression is always positive ($H^{-1}BH^{-1} \leq 0$ and $\alpha > 0$), which means that for any regularization B the covariance will be reduced to the non-regularized case. Let's summarize the results for the PR-method applied to the signal power prior case. For $B = 1$ this is equivalent to the Tikhonov regularization.

$$\boxed{\begin{aligned} \langle \mathbf{x} \rangle &= \mathbf{x}_0 \\ \mathrm{Cov}(\mathbf{x}) &= H^{-1} - 2\alpha H^{-1} B H^{-1} \end{aligned}} \tag{4.10}$$

4.2.4 Bayesian reconstruction via MCMC

In the MCMC-approach the reconstructions are sampled from the posterior distribution which has been formulated in the Bayesian interpretation of the problem. That can be done for any model with any regularization method. Here we will apply it to SENSE using a signal power prior regularization with cost function Φ in equation 3.6. To make the notation consistent and comparable with the one used for the PR-method $\mathbf{y}$ is replaced with $\mathbf{y}_0$ which is the single measurement that is going to be reconstructed. The posterior has the form

$$\begin{aligned} p(\mathbf{x}|\mathbf{y}_0, C, \mathcal{I}) &= \frac{1}{Z} p(\mathbf{y}_0|\mathbf{x}, C, \mathcal{I}) p(\mathbf{x}|C, \mathcal{I}) \\ &= \frac{1}{Z'} e^{-\frac{1}{2}\Phi(\mathbf{x})} \\ \Phi(\mathbf{x}) &:= \big(\mathbf{y}_0 - \mathcal{A}\mathbf{x}\big)^\dagger C^{-1} \big(\mathbf{y}_0 - \mathcal{A}\mathbf{x}\big) + 2\alpha \mathbf{x}^\dagger B \mathbf{x} \end{aligned} \tag{4.11}$$

The above expression uses data $\mathbf{y}_0$, noise covariance C and the model $\mathcal{A}$ as well as the regularization matrix B. An efficient numerical evaluation is presented in the appendix (see 6.2). However, to calculate mean and covariance analytically it has to be reformulated using the explicit solution of the problem in the form of the matrix H. With $H = \mathcal{A}^\dagger C^{-1} \mathcal{A} + 2\alpha B$ and $H\mathbf{x}_0 = \mathcal{A}^\dagger C^{-1} \mathbf{y}_0$ the cost function can be written as

$$\begin{aligned} \Phi(\mathbf{x}) &= \mathbf{y}_0^\dagger C^{-1} \mathbf{y}_0 - \mathbf{y}_0^\dagger C^{-1} \mathcal{A}\mathbf{x} - \mathbf{x}^\dagger \mathcal{A}^\dagger C^{-1} \mathbf{y}_0 + \mathbf{x}^\dagger \mathcal{A}^\dagger C^{-1} \mathcal{A}\mathbf{x} + 2\alpha \mathbf{x}^\dagger B \mathbf{x} \\ &= \mathbf{y}_0^\dagger C^{-1} \mathbf{y}_0 - \mathbf{x}_0^\dagger H \mathbf{x} - \mathbf{x}^\dagger H \mathbf{x}_0 + \mathbf{x}^\dagger H \mathbf{x} \\ &= \mathbf{y}_0^\dagger C^{-1} \mathbf{y}_0 - \mathbf{x}_0^\dagger H \mathbf{x} + \mathbf{x}^\dagger H (\mathbf{x} - \mathbf{x}_0) \end{aligned}$$

Adding and subtracting $\mathbf{x}_0^\dagger H \mathbf{x}_0$ and $\Delta\mathbf{x} := \mathbf{x} - \mathbf{x}_0$ leads to

$$\begin{aligned} \Phi(\mathbf{x}) &= \mathbf{y}_0^\dagger C^{-1} \mathbf{y}_0 + (\mathbf{x} - \mathbf{x}_0)^\dagger H (\mathbf{x} - \mathbf{x}_0) - \mathbf{x}_0^\dagger H \mathbf{x}_0 \\ &= \mathbf{y}_0^\dagger C^{-1} \mathbf{y}_0 + \Delta\mathbf{x}^\dagger H \Delta\mathbf{x} - \mathbf{y}_0^\dagger C^{-1} \mathcal{A} H^{-1} \mathcal{A}^\dagger C^{-1} \mathbf{y}_0 \end{aligned}$$

Terms depending on $\mathbf{y}_0$ can be absorbed in a constant when plugging Φ back into the PDF. Then the posterior PDF for the Bayesian reconstruction with signal power prior (Tikhonov) regularization is a multivariate normal distribution with the following properties:

$$\boxed{\begin{aligned} p(\mathbf{x}|\mathbf{y}_0, C, \mathcal{I}) &= \frac{1}{Z''} e^{-\frac{1}{2}\Delta\mathbf{x}^\dagger H \Delta\mathbf{x}} \\ \langle \mathbf{x} \rangle &= \mathbf{x}_0 \\ \mathrm{Cov}(\mathbf{x}) := \langle \Delta\mathbf{x}^\dagger \Delta\mathbf{x} \rangle &= H^{-1} \end{aligned}} \tag{4.12}$$

That shows that the PR method and the Bayesian method have the same mean but different covariance. In particular the covariance of the Bayesian method is generally by $H^{-1}BH^{-1}$ greater than the one of the PR method. Only for $B = 0$ (no regularization) they are equivalent.

4.3 Noise statistics in image space

The essential quantity of all the quality measures presented in this thesis is the covariance matrix of noise in image space. Pixels of MR images are complex numbers in principle but in many applications their magnitude is displayed as the "MR image". In a clinical situation a radiologist would base his diagnosis on such a magnitude image. For ROIs where he is not sure about the reliability of structures he sees, a quality measure - be it a scalar number or a quality map - can be helpful. It is important here to quantify the quality of the magnitude image rather than the complex because that is what the radiologist sees. Since the transformation from complex to magnitude data is nonlinear, also the covariance matrix changes. For single, uncorrelated pixels in MRI a derivation for the correct expression of the noise statistics for magnitude data has been given by Gudbjartsson in [10]. It turns out to be a Rician distribution if real and imaginary part of the reconstructed images are Gaussian distributed with identical covariance. In the following the the derivation for correlated pixels - which is a generalization of [10] - is given.

4.3.1 Complex multivariate Gaussian random variables

Let N complex valued pixels be assembled to a vector

$$\mathbf{z} = \mathbf{x} + i\mathbf{y}$$

in which the j-th pixel is a complex number $z_j = x_j + iy_j$. Real and imaginary parts of $\mathbf{z}$ are both multivariate Gaussian with means $\mathbf{x}^{(0)}$ and $\mathbf{y}^{(0)}$, uncorrelated with each other and have the same covariance matrix $C \in \mathrm{R}$. Then the joint PDF is the product of the PDFs for $\mathbf{x}$ and $\mathbf{y}$ and reads

$$\begin{aligned} p(\mathbf{x}, \mathbf{y}|\mathbf{x}_0, \mathbf{y}_0) &= \frac{1}{Z}\exp\Big(-\frac{\phi}{2}\Big) \\ \phi &= \Delta\mathbf{x}^T C^{-1}\Delta\mathbf{x} + \Delta\mathbf{y}^T C^{-1}\Delta\mathbf{y} \\ \Delta\mathbf{x} &= \mathbf{x} - \mathbf{x}^{(0)} \\ \Delta\mathbf{y} &= \mathbf{y} - \mathbf{y}^{(0)} \\ Z &= (2\pi)^N |C| \end{aligned}$$

To evaluate mean and covariance of the complex multivariate Gaussian we are interested in the marginal PDFs of

1. a single pixel z_i (to calculate mean and variance) $\longrightarrow N_0 = 1$
2. two pixels (z_i, z_j) (for the covariance matrix) $\longrightarrow N_0 = 2$

The evaluation of both cases can be simplified by renumbering the pixels of $\mathbf{z}$ such that the inverse covariance matrix becomes an $(N_0N_1 \times N_0N_1)$ block structure for each z_i or each pair of (z_i, z_j). The switch for above's cases N_0 is either 1 or 2 and $N_1 = N - N_0$. The inverse covariance matrix is then written in block form as

$$C^{-1} = \begin{pmatrix} A_0 & B \\ B^T & A_1 \end{pmatrix}$$

A_0 is the upper left $(N_0 \times N_0)$ block, A_1 is the lower right $(N_1 \times N_1)$ block and B is the remaining off-diagonal block.

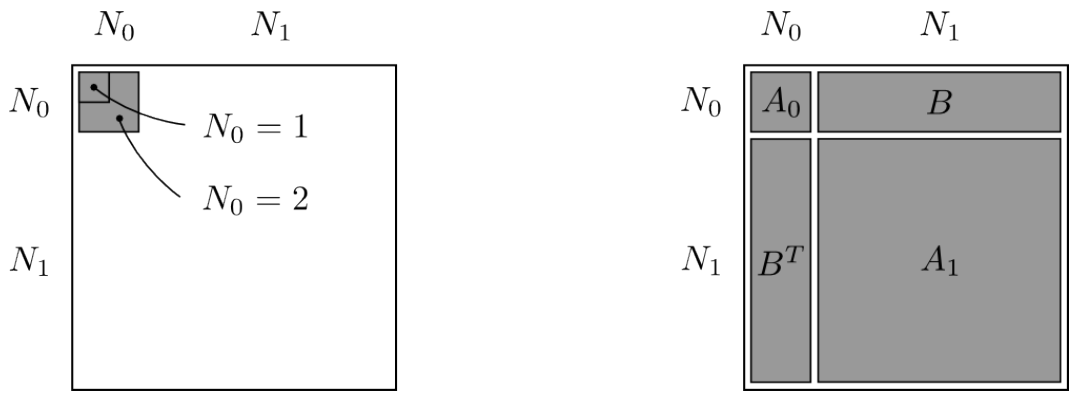


Figure 16: Renumbering of C^{-1} for each pixel ($N_0 = 1$) or each pair of pixels ($N_0 = 2$) (left), Block form of inverse covariance matrix (right).

The block matrix can be transformed to a block diagonal matrix with the help of the Schur complement matrix S and the LDU decomposition. The fact that C is a covariance matrix, hence $C = C^T$, and the general identity of $\left(M^T\right)^{-1} = \left(M^{-1}\right)^T$ have been used.

$$C^{-1} = \begin{pmatrix} A_0 & B \\ B^T & A_1 \end{pmatrix} = \begin{pmatrix} \mathbf{1} & -F \\ 0 & \mathbf{1} \end{pmatrix} \begin{pmatrix} S & 0 \\ 0 & A_1 \end{pmatrix} \begin{pmatrix} \mathbf{1} & 0 \\ -F^T & \mathbf{1} \end{pmatrix} = V^T D V$$

$$\begin{aligned} F &= -BA_1^{-1} \\ S &= A_0 - BA_1^{-1}B^T \\ V &= \begin{pmatrix} \mathbf{1} & 0 \\ -F^T & \mathbf{1} \end{pmatrix} \end{aligned}$$

The the covariance matrix itself can be written as

$$\begin{aligned} C = \begin{pmatrix} A_0 & B \\ B^T & A_1 \end{pmatrix}^{-1} &= V^{-1} D^{-1} \left(V^{-1}\right)^T \\ &= \begin{pmatrix} \mathbf{1} & 0 \\ F^T & \mathbf{1} \end{pmatrix} \begin{pmatrix} S^{-1} & 0 \\ 0 & A_1^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{1} & F \\ 0 & \mathbf{1} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{1} & 0 \\ F^T & \mathbf{1} \end{pmatrix} \begin{pmatrix} S^{-1} & S^{-1}F \\ 0 & A_1^{-1} \end{pmatrix} \\ &= \begin{pmatrix} S^{-1} & S^{-1}F \\ F^T S^{-1} & F^T S^{-1} F + A_1^{-1} \end{pmatrix} \end{aligned}$$

The desired PDF is either $p(z_i)$ or $p((z_i, z_j))$. It will be denoted as $p(\mathbf{z}_0)$, where subscript-0 variables are going to refer to the according case of either a single pixel ($\mathbf{z}_0 = z_j = \text{scalar}$) or a pair ($\mathbf{z}_0 = (z_i, z_j) = \text{2D-vector}$). By splitting the whole vector $\mathbf{z}$ into two parts $(\mathbf{z}_0, \mathbf{z}_1)$, the PDF can be computed by marginalizing over the remaining N_1 variables in $\mathbf{z}_1$. Real and imaginary parts $\mathbf{x}_1$ and $\mathbf{y}_1$ are uncorrelated and can therefore be split and both calculated in the same way. For the real part the PDF gives

$$\begin{aligned} p(\mathbf{x}_0) &\propto \int \exp\big(-\Delta\mathbf{x}^T C^-1 \Delta\mathbf{x}\big) d^{N_1} x_1 \\ &\propto \int \exp\big(-(V\Delta\mathbf{x})^T D V \Delta\mathbf{x}\big) d^{N_1} x_1 \end{aligned}$$

With the variable transform $V : \Delta\mathbf{x} \to \Delta\tilde{\mathbf{x}}$ that only affects the subscript-0 part the marginal PDF gets

$$p(\mathbf{x}_0) \propto \int \exp\big(-\Delta\tilde{\mathbf{x}}^T D \Delta\tilde{\mathbf{x}}\big) d^{N_1}\tilde{x}_1 \;\propto \exp\big(-\Delta\mathbf{x}_0^T S \Delta\mathbf{x}_0\big)$$

$$\Delta\tilde{\mathbf{x}} := V\Delta\mathbf{x} = \begin{pmatrix} \mathbf{1} & 0 \\ -F^T & \mathbf{1} \end{pmatrix} \begin{pmatrix} \Delta\mathbf{x}_0 \\ \Delta\mathbf{x}_1 \end{pmatrix} = \begin{pmatrix} \Delta\mathbf{x}_0 \\ \Delta\mathbf{x}_1 - F^T\Delta\mathbf{x}_0 \end{pmatrix}$$

because the integration only targets the N_1 variables in the lower block and $|V| = 1$. With correct normalization the mean is given by the subscript-0 part of the original mean $\mathbf{x}^{(0)}$ - denoted as $\mathbf{x}_0$ - and the marginal covariance matrix $C_0 := S$ is given by the upper left $(N_0 \times N_0)$ block of the original covariance matrix C. For the imaginary part the calculation is the same such that both can be put together for the whole complex number $\mathbf{z}_0$. That yields the statistics for a single complex pixel or a pair of complex pixels.

$$\boxed{\begin{aligned} p(\mathbf{z}_0) &= \frac{1}{Z'} \exp\Big(-\frac{1}{2}\big(\Delta\mathbf{x}_0^T C_0^{-1}\Delta\mathbf{x}_0 + \Delta\mathbf{y}_0^T C_0^{-1}\Delta\mathbf{y}_0\big)\Big) \\ Z' &= (2\pi)^{N_0}|C_0| \\ \langle\mathbf{z}_0\rangle &= \mathbf{x}_0 + i\mathbf{y}_0 \\ \mathrm{Cov}(\mathbf{z}_0) &= C_0 \end{aligned}} \tag{4.13}$$

4.3.2 Mean and variance of the magnitude of a single pixel

From this starting point the statistics for the magnitudes can be derived. As a start mean and variance of a single pixel j (case 1) can be calculated. For the complex pixel value z_j the real and complex means are given by $x_{0,j} = \mathbf{x}_j^{(0)}$ and $y_{0,j} = \mathbf{y}_j^{(0)}$ and the variance is $\sigma_j^2 = C_{jj}$. The index j will be suppressed from now onward. We are looking for the PDF for the magnitude $r := |z| = \sqrt{x^2 + y^2}$. The PDF for complex pixels above can be marginalized over all the real and imaginary parts by using a Dirac-delta for the magnitude.

$$p(r) = \frac{1}{Z'} \int \delta\big(r - \sqrt{x'^2 + y'^2}\big) \exp\Big(-\frac{1}{2\sigma^2}\big((x' - x_0)^2 + (y' - y_0)\big)^2\Big) dx'dy'$$

Now all variables are expressed in units of σ such that $x' = \sigma\xi'$, $y' = \sigma\eta'$, $x_0 = \sigma\xi_0$, $y_0 = \sigma\eta_0$ and $r = \sigma\rho$.

$$\begin{aligned} p(r) &= \frac{1}{2\pi} \int \delta\big(\sigma(\rho - \sqrt{\xi'^2 + \eta'^2})\big) \exp\Big(-\frac{1}{2}\big((\xi' - \xi_0)^2 + (\eta' - \eta_0)\big)^2\Big) d\xi'd\eta' \\ &= \frac{1}{2\pi\sigma}\, I(\rho|\xi_0, \eta_0) \\ I(\rho|\xi_0, \eta_0) &:= \int \delta\big(\rho - \sqrt{\xi'^2 + \eta'^2}\big) \exp\Big(-\frac{1}{2}\big(\xi'^2 + \eta'^2 + \xi_0^2 + \eta_0^2 - 2\big(\xi'\xi_0 + \eta'\eta_0\big)\Big) d\xi'd\eta' \end{aligned}$$

The fact that for a single pixel $Z' = 2\pi\sigma^2$ and $\delta(\sigma x) = \delta(x)/|\sigma|$ have been used. The pairs (ξ', η') and (ξ_0, η_0) can be considered as vectors of length ρ' and ρ_0. Using polar coordinates

(ρ', φ') the integral is written as

$$
\begin{aligned}
I(\rho|\xi_0, \eta_0) &:= \int_0^\infty \rho' d\rho' \int_0^{2\pi} \delta\big(\sigma(\rho - \rho')\big)\, \exp\Big(-\frac{1}{2}\big(\rho'^2 + \rho_0^2 - 2\rho'\rho_0 \cos\varphi'\big)\Big) d\varphi' \\
&= \rho e^{-\frac{\rho^2+\rho_0^2}{2}} \int_0^{2\pi} \exp\Big(\rho\rho_0 \cos\varphi'\Big) d\varphi'
\end{aligned}
$$

Here the integral representation of the modified Bessel function $I_0\big(z\big) := \frac{1}{\pi}\int_0^\pi \exp\big(\pm z\cos\varphi'\big) d\varphi'$ is used. The integration from $[0, 2\pi]$ is equal to $[-\pi, \pi]$.

$$
\begin{aligned}
I(\rho|\xi_0, \eta_0) &= \rho e^{-\frac{\rho^2+\rho_0^2}{2}} \int_0^{2\pi} \exp\Big(\rho\rho_0 \cos\varphi'\Big) d\varphi' \\
&= \rho e^{-\frac{\rho^2+\rho_0^2}{2}} \int_{-\pi}^{\pi} \exp\Big(\rho\rho_0 \cos\varphi'\Big) d\varphi' \\
&= 2\pi\rho e^{-\frac{\rho^2+\rho_0^2}{2}} I_0\big(\rho\rho_0\big)
\end{aligned}
$$

The result is the Rician distribution for the magnitude of single pixels.

$$
\boxed{
\begin{aligned}
p(r) &= \frac{r}{\sigma^2} e^{-\frac{r^2+r_0^2}{2\sigma^2}} I_0\Big(\frac{r r_0}{\sigma^2}\Big) \\
\sigma^2 &= C_{jj} \\
r &= |z_j| \\
r_0 &= \big|\mathbf{z}_j^{(0)}\big|
\end{aligned}
} \tag{4.14}
$$

4.3.3 Covariance of the magnitudes of two pixels

The full covariance matrix for the magnitudes of complex multivariate Gaussian random variables can be computed element by element from the joint PDF for two pixels which can expressed by marginalizing over all pixels except the ones in question. The renumbering of the original matrix was performed in such a way that the pixels (i, j) get $i \to 1$ and $j \to 2$. Then for any pixel pair we get

$$
\begin{aligned}
p(r_1, r_2) &\propto \int \delta(r_1 - |z_1'|)\delta(r_2 - |z_2'|)\, \exp\Big(-\frac{1}{2\sigma^2}\Phi\Big) dz_1' dz_2' \\
\Phi &:= \Delta\mathbf{x}^T A \Delta\mathbf{x} + \Delta\mathbf{y}^T A \Delta\mathbf{y} \\
A &:= C_0^{-1} = \begin{pmatrix} C(i,i) & C(i,j) \\ C(j,i) & C(j,j) \end{pmatrix}
\end{aligned}
$$

where Φ is rewritten as

$$\begin{aligned}
\Phi &:= \Delta\mathbf{x}^T A \Delta\mathbf{x} + \Delta\mathbf{y}^T A \Delta\mathbf{y} \\
&= \mathbf{x}^T A \mathbf{x} + \mathbf{y}^T A \mathbf{y} + \mathbf{x}_0^T A \mathbf{x}_0 + \mathbf{y}_0^T A \mathbf{y}_0 - 2\mathbf{x}^T A \mathbf{x}_0 - 2\mathbf{y}^T A \mathbf{y}_0 \\
&= (x_1^2 + y_1^2)A_{11} + (x_{10}^2 + y_{10}^2)A_{11} + (x_2^2 + y_2^2)A_{22} + (x_{20}^2 + y_{20}^2)A_{11} + 2(x_1x_2 + y_1y_1)A_{12} + \\
&\quad + 2(x_{10}x_{20} + y_{10}y_{10})A_{12} - 2(x_1x_{10} + y_1y_{10})A_{11} - 2(x_2x_{20} + y_2y_{20})A_{22} - 2(x_1x_{20} + y_1y_{20})A_{12} - \\
&\quad - 2(x_2x_{10} + y_2y_{10})A_{12} \\
&= (r_1^2 + r_{10}^2)A_{11} + (r_2^2 + r_{20}^2)A_{22} + 2\mathbf{z}_1^T\mathbf{z}_2 A_{12} + 2\mathbf{z}_{10}^T\mathbf{z}_{20}A_{12} - 2\mathbf{z}_1^T\mathbf{z}_{10}A_{11} - 2\mathbf{z}_2^T\mathbf{z}_{20}A_{22} - \\
&\quad - 2\mathbf{z}_1^T\mathbf{z}_{20}A_{12} - 2\mathbf{z}_2^T\mathbf{z}_{10}A_{12}
\end{aligned}$$

The vector $\mathbf{z}_\alpha$ is defined as $\mathbf{z}_\alpha := \begin{pmatrix} x_\alpha \\ y_\alpha \end{pmatrix}$. Collecting terms that are independent of $\mathbf{z}_1$ and $\mathbf{z}_2$ in a constant $\tilde{C}$ and separating terms depending on $\mathbf{z}_1$ to perform the integral over pixel 1 analytically yields

$$\Phi = \tilde{C} + r_1^2 A_{11} + r_2^2 A_{22} + 2\mathbf{z}_1^T\big[\mathbf{z}_2 A_{12} - \underbrace{(\mathbf{z}_{10}A_{11} + \mathbf{z}_{20}A_{12})}_{=:\mathbf{c}}\big] - 2\mathbf{z}_2^T \underbrace{[\mathbf{z}_{20}A_{22} + \mathbf{z}_{10}A_{12}]}_{=:\mathbf{d}}$$

The PDF gets then

$$p(r_1, r_2) \propto e^{-\frac{1}{2}(A_{11}r_1^2 + A_{22}r_2^2)} \int dz_2' \; \delta(r_2 - |z_2'|) \exp\Big(\mathbf{z}_2'^T\mathbf{d}\Big) \underbrace{\int dz_1' \; \delta(r_1 - |z_1'|) \exp\Big(- \mathbf{z}_1'^T(\mathbf{z}_2' A_{12} - \mathbf{c})\Big)}_{=:T}$$

The integral over T can be evaluated by transforming $\mathbf{z}_1'$ into polar coordinates:

$$\begin{aligned}
T &= r_1 \int_{-\pi}^{\pi} d\varphi' \; \exp\big(- r_1|\mathbf{b}|\cos\varphi'\big) = 2\pi r_1 I_0(r_1|\mathbf{b}|) \\
\mathbf{b} &:= \mathbf{z}_2' A_{12} - \mathbf{c} \\
|\mathbf{b}| &= \sqrt{A_{12}^2 r_2^2 + c^2 - 2A_{12}\mathbf{z}_2'^T\mathbf{c}}
\end{aligned}$$

The PDF simplifies to

$$p(r_1, r_2) \propto r_1 e^{-\frac{1}{2}(A_{11}r_1^2 + A_{22}r_2^2)} \int dz_2' \; \delta(r_2 - |z_2'|) \exp\Big(\mathbf{z}_2'^T\mathbf{d}\Big) I_0(r_1|\mathbf{b}|)$$

The remaining integral can be simplified a bit by modifying the two scalar products $\mathbf{z}_2'^T\mathbf{c}$ and $\mathbf{z}_2'^T\mathbf{d}$. The direction of $\mathbf{d}$ can be written as a unit vector $\mathbf{e}$

$$\mathbf{e} := \frac{\mathbf{d}}{|\mathbf{d}|} = \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}$$

A second unit vector, orthogonal to $\mathbf{e}$, is defined as

$$\mathbf{e}_\perp := \begin{pmatrix} -e_2 \\ e_1 \end{pmatrix}$$

Embedding these vectors in R^3 by adding a 0 as the third component enables us to form a vector product

$$\mathbf{e} \times \mathbf{e}_\perp = \begin{pmatrix} e_1 \\ e_2 \\ 0 \end{pmatrix} \times \begin{pmatrix} -e_2 \\ e_1 \\ 0 \end{pmatrix} = \mathbf{e}_z$$

After introducing an integration angle φ' and expanding $\mathbf{c}$ in terms of the two unit vectors, the scalar products can be written explicitly as

$$\begin{aligned}
\mathbf{z}_2'^T\mathbf{d} &= r_2\ d\ \cos\varphi' \\
\mathbf{c} &= \underbrace{\left(\mathbf{d}^T\mathbf{e}\right)}_{c_1}\mathbf{e} + \underbrace{\left(\mathbf{d}^T\mathbf{e}_\perp\right)}_{c_2}\mathbf{e}_\perp \\
\mathbf{z}_2'^T\mathbf{c} &= c_1\mathbf{z}_2'^T\mathbf{e} + c_2\mathbf{z}_2'^T\mathbf{e}_\perp = r_2(c_1\ \cos\varphi' + c_2\ \sin\varphi')
\end{aligned}$$

The integral in polar coordinates collapses in r_2' due to the Delta. Finally the remaining integral - from which the non diagonal elements of the Rician covariance matrix can be computed - becomes

$$\begin{aligned}
p(r_1, r_2) &\propto r_1 r_2 e^{-\frac{1}{2}(A_{11}r_1^2 + A_{22}r_2^2)} \int_0^{2\pi} d\varphi'\ \exp\left(r_2 d\ \cos\varphi'\right) I_0(r_1 b(\varphi')) \\
b(\varphi') &= \sqrt{A_{12}^2 r_2^2 + c^2 - 2A_{12}\mathbf{z}_2'^T\mathbf{c}} \\
&= \sqrt{A_{12}^2 r_2^2 + c^2 - 2A_{12}r_2(c_1\cos\varphi' + c_2\sin\varphi')} \\
&= \sqrt{\alpha + \beta_1\cos\varphi' + \beta_2\sin\varphi')} \\
\alpha &= A_{12}^2 r_2^2 + c^2 \\
\beta_1 &= -2A_{12}r_2 c_1 \\
\beta_2 &= -2A_{12}r_2 c_2 \\
\mathbf{c} &= \mathbf{z}_{10}A_{11} + \mathbf{z}_{20}A_{12} \\
\mathbf{d} &= \mathbf{z}_{20}A_{22} + \mathbf{z}_{10}A_{12} \\
\mathbf{e} &= \frac{\mathbf{d}}{|\mathbf{d}|} = \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \\
\mathbf{e}_\perp &= \begin{pmatrix} -e_2 \\ e_1 \end{pmatrix} \\
c_1 &= \mathbf{d}^T\mathbf{e} \\
c_2 &= \mathbf{d}^T\mathbf{e}_\perp
\end{aligned} \tag{4.15}$$

As a test this general expression can be compared to the expression for the variance ($A_{12} = 0$). In this case the constants become

$$\begin{aligned}
b(\varphi') &= A_{11}r_{10} \\
\alpha &= c^2 = r_{10}^2 A_{11} \\
\beta_1 &= \beta_2 = 0 \\
d &= r_{20} \\
p(r_1, r_2) &\propto r_1 r_2\ e^{-\frac{1}{2}(A_{11}r_1^2 + A_{22}r_2^2)}\ I_0(r_1 r_{10} A_{11})\ I_0(r_2 r_{20} A_{22}) \\
&\propto \frac{r_1}{\sigma_{11}^2} e^{-\frac{r_1^2 + r_{10}^2}{2\sigma_{11}^2}} I_0\left(\frac{r_1 r_{10}}{\sigma_{11}^2}\right) \frac{r_2}{\sigma_{22}^2} e^{-\frac{r_2^2 + r_{20}^2}{2\sigma_{22}^2}} I_0\left(\frac{r_2 r_{20}}{\sigma_{22}^2}\right) = p(r_1)\ p(r_2)
\end{aligned}$$

This is the expected result for the marginal PDF of two uncorrelated pixels, formed by the product of the PDFs for the two single pixels. The evaluation of the remaining integral over φ in order to compute the 2×2 covariance matrix has to be done numerically.

4.4 Statistical image quality measures

4.4.1 G-factor maps

As defined in a paper by Pruessmann in [19], geometry factor maps are a quality measure of MR image reconstructions from undersampled data. The idea is basically to compare the sample reconstruction (of which the quality needs to be estimated) with the reconstruction of the fully sampled dataset. The latter therefore serves as a relative "gold standard". Precisely the g-factor map of a reconstruction is given by the fraction of the pixel-wise SNR in the image domain from the gold standard and the sample. The result is a map of SNR amplification with respect to the fully sampled image. The SNR in image space can be deduced analytically for certain problems (see 4.2.3) or obtained by the PR method (see 4.2.1) or an MCMC run (see 4.2.4). The acceleration factor acc is defined as the fraction of the full number of k-space points and the reduced number when undersampling.

$$\begin{aligned} g(\mathbf{x}, acc) &:= \frac{1}{\sqrt{acc}} \frac{SNR_{full}(\mathbf{x})}{SNR_{acc}(\mathbf{x})} = \frac{1}{\sqrt{acc}} \frac{\frac{\langle \mathbf{x}_{full} \rangle}{\sigma_{full}(\mathbf{x})}}{\frac{\langle \mathbf{x}_{acc} \rangle}{\sigma_{acc}(\mathbf{x})}} \qquad \text{with} \quad g \geq 1 \\ acc &:= \frac{N_{\mathbf{y},full}}{N_{\mathbf{y},acc}} \end{aligned} \tag{4.16}$$

Hence g-factor maps represent a pixel-wise quality measure with respect to a fully sampled gold standard. The higher g is in a certain pixel (or ROI), the worse is the quality. However this quality measure does not take pixel correlations into account.

4.4.2 The covariance quality measure

A general way to answer the question whether a certain structure in given data is "real" or due to noise is to use the covariance quality measure. It does not only take into account the STD of the pixel but also considers correlations between them. Therefore it is a generalization of SNR used in g-factor maps. To derive a covariance matrix of reconstructed images either the PR-method or a MCMC-reconstruction has to be applied. That means that the pixel values are interpreted as random variables. Given a certain number of positive pixel values x_i from a certain region I of the image (e.g. the magnitude), a new random variable F can be defined as the integral over these pixels.

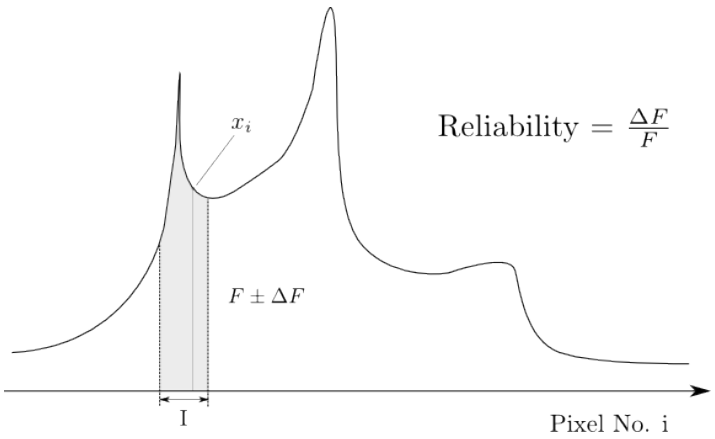


Figure 17: Covariance quality measure for a region I of the image to differentiate between "real" and noisy structures.

The area under the curve can be approximated by a sum over the pixel values, setting $\Delta i = 1$, and we arrive at the following expressions:

$$\begin{aligned} F &:= \sum_{i \in I} x_i \\ \langle F \rangle &= \langle \sum_{i \in I} x_i \rangle = \sum_{i \in I} \langle x_i \rangle \\ \mathrm{Cov}(F) &= \langle (F - \langle F \rangle)^2 \rangle = \Big\langle \Big(\sum_{i \in I} x_i - \sum_{i \in I} \langle x_i \rangle \Big)^2 \Big\rangle = \Big\langle \Big(\sum_{i \in I} \underbrace{x_i - \langle x_i \rangle}_{\Delta x_i} \Big)^2 \Big\rangle \\ &= \sum_{i,j \in I} \langle \Delta x_i \Delta x_j \rangle = \sum_{i,j \in I} \mathrm{Cov}(x_i, x_j) \end{aligned} \tag{4.17}$$

It is intuitive now to assume that the structure in region I is of "real" origin when $\Delta F = \sqrt{\mathrm{Cov}(F)}$ is of orders smaller than F. Otherwise it is very likely that the structure I we see is due to noise and must be regarded as unreliable. A special case of equation 4.17 is the case where all pixels are uncorrelated. Then only the variances contribute to F.

$$Quality = \begin{cases} good, & \text{if } \frac{\Delta F}{F} << 1 \\ bad, & \text{if } \frac{\Delta F}{F} \approx 1 \text{ or } > 1 \end{cases} \tag{4.18}$$

5 Results and Discussion

5.1 Noise statistics in k-space

The noise of the signal in k-space is determined by measuring noise samples in the absence of magnetic gradients (no spatial encoding) with n_c parallel receiver coils. That is equivalent to taking measurements in the k-space center only, because for $\mathbf{k} = \mathbf{0}$ the signal equation collapses. Any k-space signal captured by the receiver coils is encoded by the scanner as a complex number like $n = a + bi$, so the noise samples are complex $n_c \times 1$ vectors: one data point at k-space center per coil. That is why we the scanner returns a $n_c \times n_c$ coil covariance matrix. Ideally the coils are independent from each other so the coil covariance matrix for the complex noise samples n should be diagonal. For further considerations it is useful to split the noise samples $\{n\}$ into real and imaginary part and analyze the covariance matrix of the real and imaginary part on their own. Doing so it turns out that $\mathrm{Cov}(a)$ and $\mathrm{Cov}(b)$ are not only diagonal but also identical. In figure 18 the covariance matrices $\mathrm{Cov}(a)$ and $\mathrm{Cov}(b)$ of the real and imaginary part of a noise sample are shown. The fact that $\mathrm{Cov}(a) \approx \mathrm{Cov}(b)$ is underlined by the histograms in figure 19 and 20 and they show that real and imaginary part of the k-space noise is Gaussian.

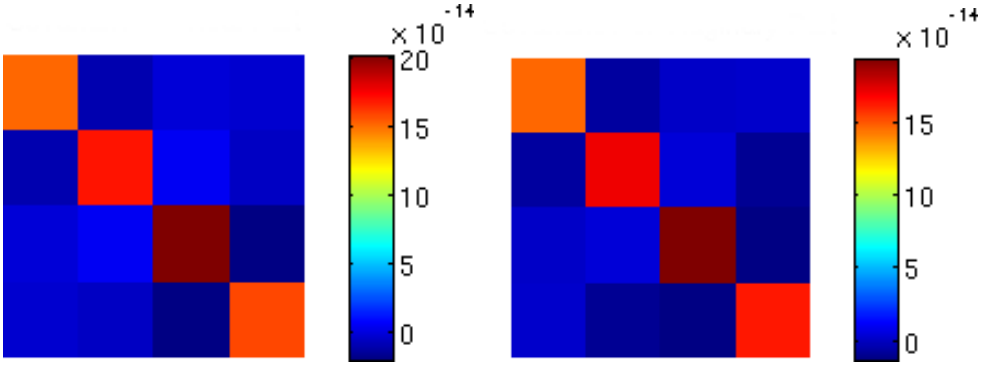


Figure 18: Covariance matrices $\mathrm{Cov}(a)$ (left) and $\mathrm{Cov}(b)$ (right) of real and imaginary part of the complex noise samples $n = a + ib$. It shows that they are identical for the real and imaginary part. Also the $n_c = 4$ coils are relatively independent.

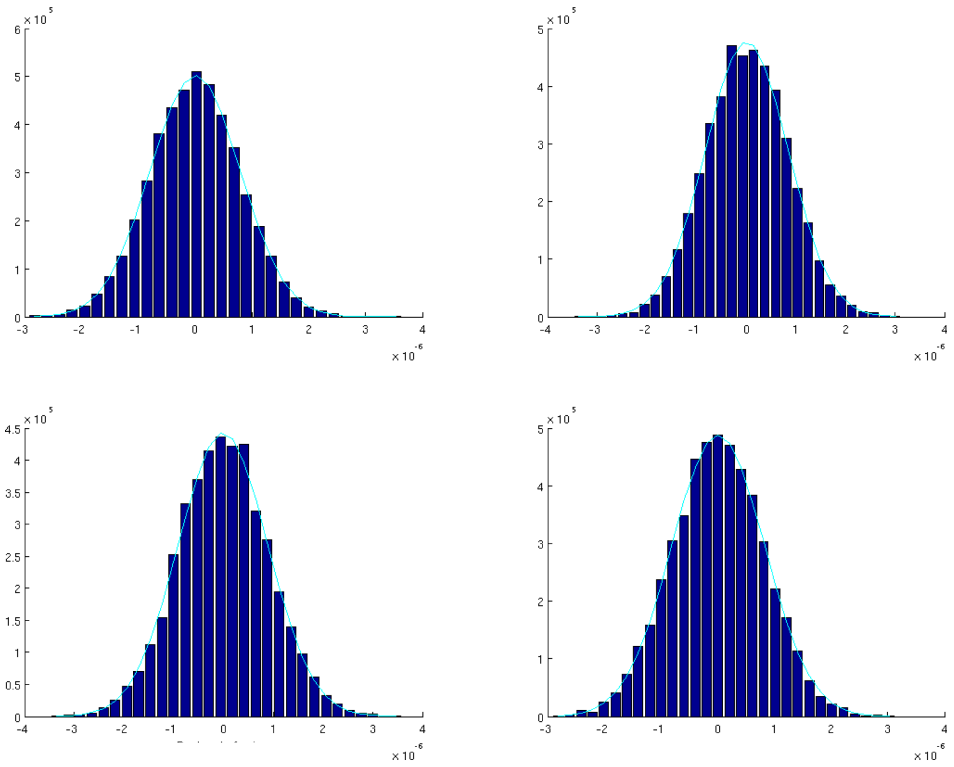


Figure 19: Histogram of the real part of k-space noise (in $\mathbf{k} = \mathbf{0}$) for $n_c = 4$ receiver coils.

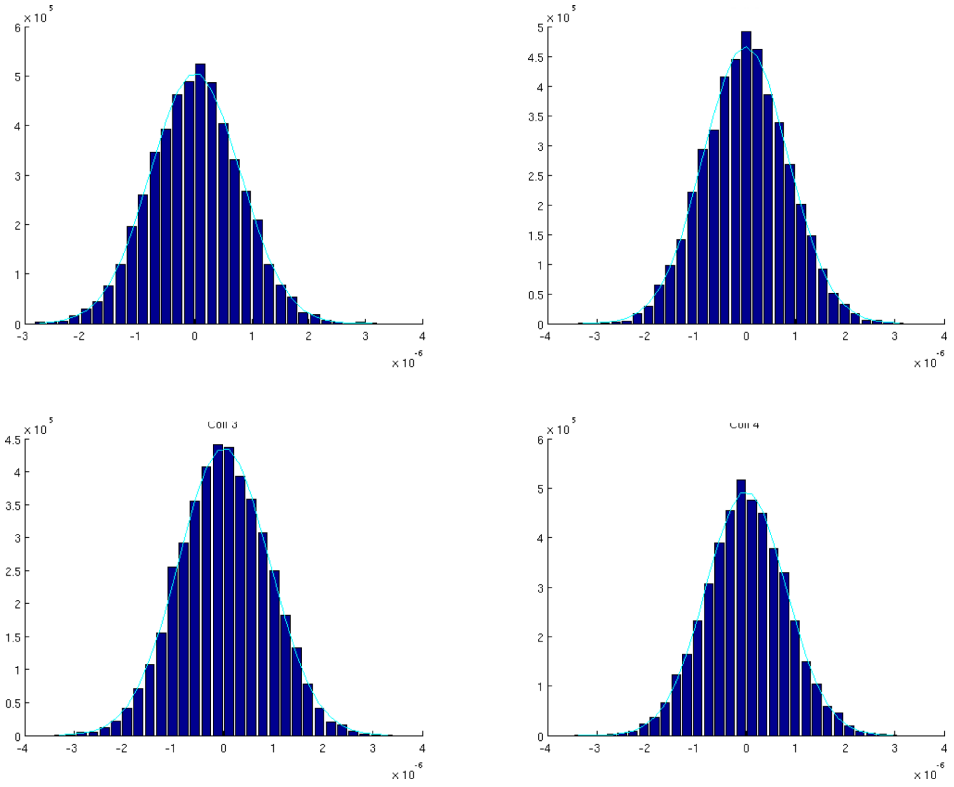


Figure 20: Histogram of the imaginary part of k-space noise (in $\mathbf{k} = \mathbf{0}$) for $n_c = 4$ receiver coils.

For covariance matrices of complex random variables $\mathbf{x} = \mathbf{a} + i\mathbf{b}$ the following statement holds if $\mathrm{Cov}(a) \approx \mathrm{Cov}(b)$:

$$\begin{aligned}\mathrm{Cov}(\mathbf{x}) &= \mathrm{Cov}(\mathbf{a}+i\mathbf{b}) = \langle(\mathbf{a}+i\mathbf{b}-\langle\mathbf{a}+i\mathbf{b}\rangle)(\mathbf{a}+i\mathbf{b}-\langle\mathbf{a}+i\mathbf{b}\rangle)^\dagger\rangle \\ &= \langle(\mathbf{a}+i\mathbf{b}-\langle\mathbf{a}\rangle - i\langle\mathbf{b}\rangle)(\mathbf{a}^\dagger - i\mathbf{b}^\dagger - \langle\mathbf{a}\rangle^\dagger + i\langle\mathbf{b}\rangle^\dagger)\rangle \\ &= \langle(\mathbf{a}-\langle\mathbf{a}\rangle)^2 + (\mathbf{b}-\langle\mathbf{b}\rangle)^2\rangle \\ &= \mathrm{Cov}(\mathbf{a}) + \mathrm{Cov}(\mathbf{b})\end{aligned} \tag{5.1}$$

These insights are very useful because they can be included in further procedures. When reconstructing an image the knowledge of the k-space noise distribution (in the form of weighted LSQ) will increase the condition of the problem because any measured data will be subject to such noise. Also for generating artificially noised data for the PR-method this information is crucial. So it is necessary to know the covariance matrix of the noise that corrupts k-space data - which is actually a huge $n_x n_y n_c \times n_x n_y n_c$ matrix. The $n_c \times n_c$ covariance matrix obtained by the scanner does not seem to fit into that concept. In fact, it is only a compressed version of the "full" matrix because every k-space point has the same noise level. That is due to the fact that noise is defined as the signal without spatial encoding (the k-space center $\mathbf{k} = \mathbf{0}$). Therefore the "full" k-space covariance matrix is formed by copies of the "small" one like shown in figure 20. Numerically it is still more convenient to rearrange the data such that the small covariance matrix can be used.

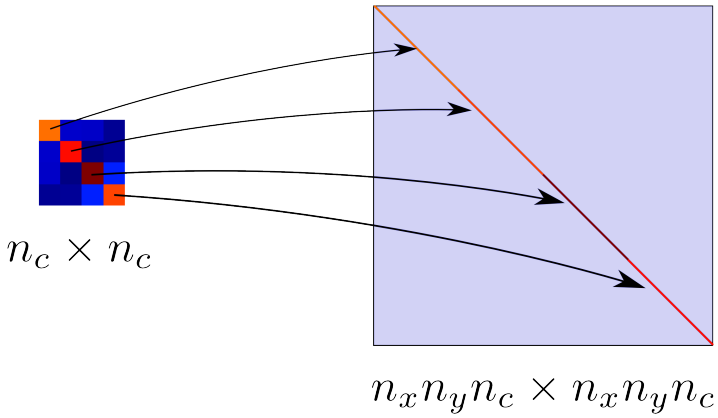


Figure 21: The full covariance matrix is formed by replicating the elements of the "small" one provided by the MR scanner. Here $n_c = 4$ coils are used and only the replication of the diagonal elements is shown.

In the following figure 22 the sparsity pattern and a detailed view of the covariance matrix of k-space noise is shown for different acceleration factors. It has been calculated using simulated data samples with additive Gaussian noise according to the 4×4 coil covariance matrix in figure 18. A data simulation process of this kind is used for the PR method (see section 4.2.1). As expected, it turns out that the $n_x n_y n_c \times n_x n_y n_c$ matrix is also diagonal. A closer look also reveals the sampling trajectory, which fully samples the k-space and leaves out every acc^{th} point in one k-space direction.

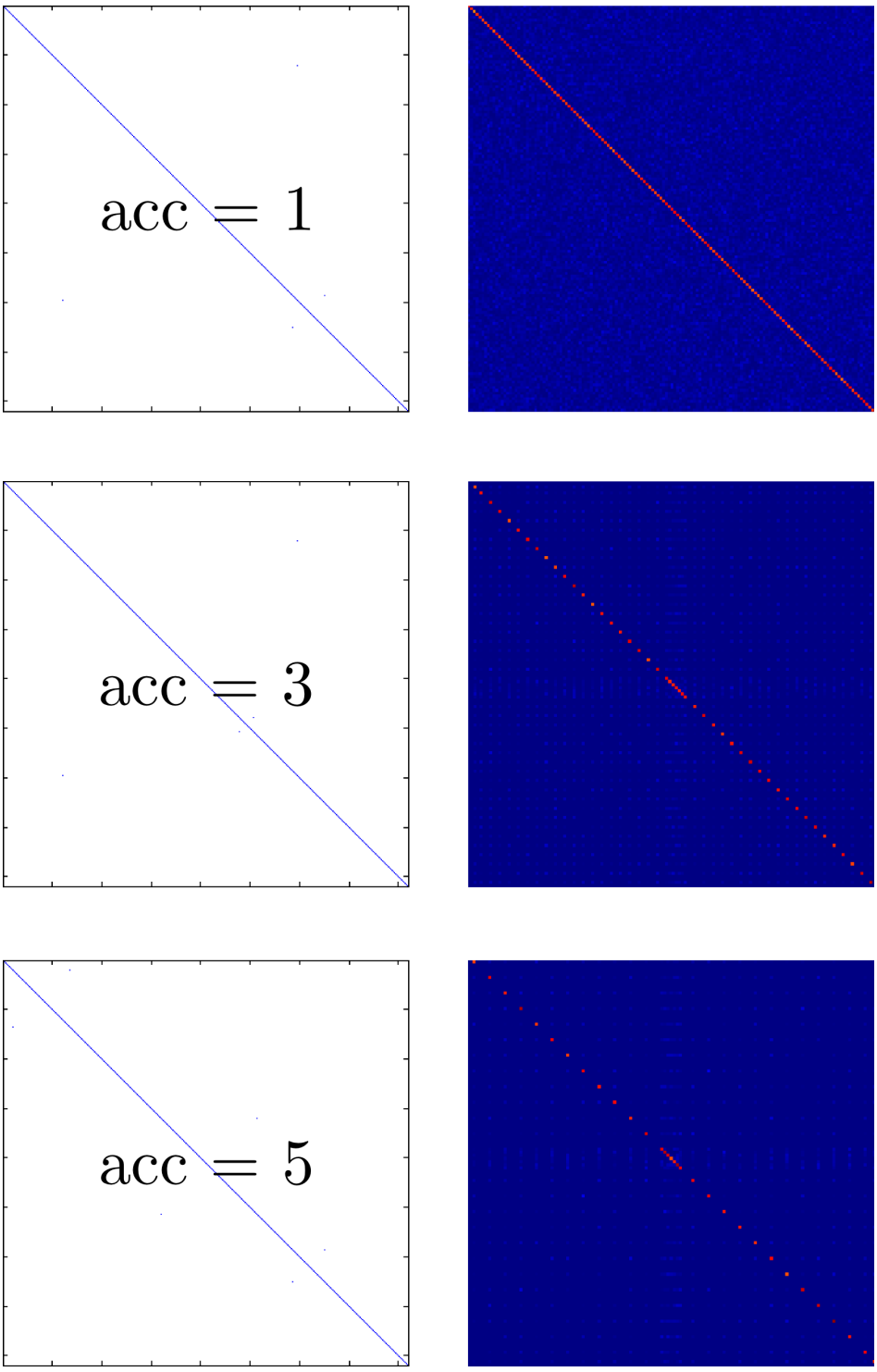


Figure 22: Covariance matrix of k-space noise: The sparsity pattern of the full $n_x n_y n_c \times n_x n_y n_c$ matrix is plotted in the first column and shows that the covariance matrix is diagonal for different accelerations (acc = 1, 2 and 5). Elements smaller than 10 % of the maximal value are neglected. In the second column a detailed view is plotted which reveals the sampling trajectory for a single k-space line. According to the acceleration factor, k-space points are left out periodically whilst sampling, only the k-space center remains fully sampled.

5.2 Model validation

A straight forward check to see whether the reconstructed image together with the model fits the data is to compute the number of degrees of freedom (DoF) via the χ^2-test and compare it with the theoretically expected number. For the following graphs a 40 x 40 dataset measured by 4 coils with Cartesian sampling trajectory have been used. For different acceleration factors each 2nd, 3rd, 4th, ... k-space line in k_x-direction was omitted for reconstruction, whilst k_y-direction was fully sampled. The k-space center was fully sampled by keeping 5% of the lines in k_x direction. The theoretical number for the DoF of the problem is known as $N = c_{\mathcal{S}} n_x n_y n_c - n_x n_y$. It is a lower bound for χ^2, because the 4 coils are not fully linear independent. Therefore the effective number of DoF will be higher. The factor $c_{\mathcal{S}}$ depends on the undersampling trajectory and is equal to 1 for a fully sampled k-space (for undersampling $c_{\mathcal{S}} < 1$). N and χ^2 decrease with enforced undersampling because less data is collected. Figure 24 shows that the model is fitting the data well for small accelerations with deviation towards higher accelerations. Theoretically it is expected that undersampling according to acceleration factor 4 should still be possible with 4 receiver coils if the coil sensitivities are linear independent. However, this is not the case in reality: For acceleration factors of 3 and higher the reconstructions in figure 24 become more and more governed by noise because the problem gets increasingly ill-conditioned. The coil sensitivities as the source of linear dependency start corrupting the result of the reconstruction, the effective number of DoF decreases. With increasing acceleration it becomes necessary to regularize in order to get reliable images. Figure 23 shows the connection between the condition number of the problem and the deviation of the effective number of DoF with respect to the theoretically expected one. Both numbers grow exponentially with increased undersampling. The right plot explains the increasing noise corruption of the reconstructed images in figure 24 for $acc > 2$: It shows that the difference in theoretical and expected DoF has a kink at $acc = 3$.

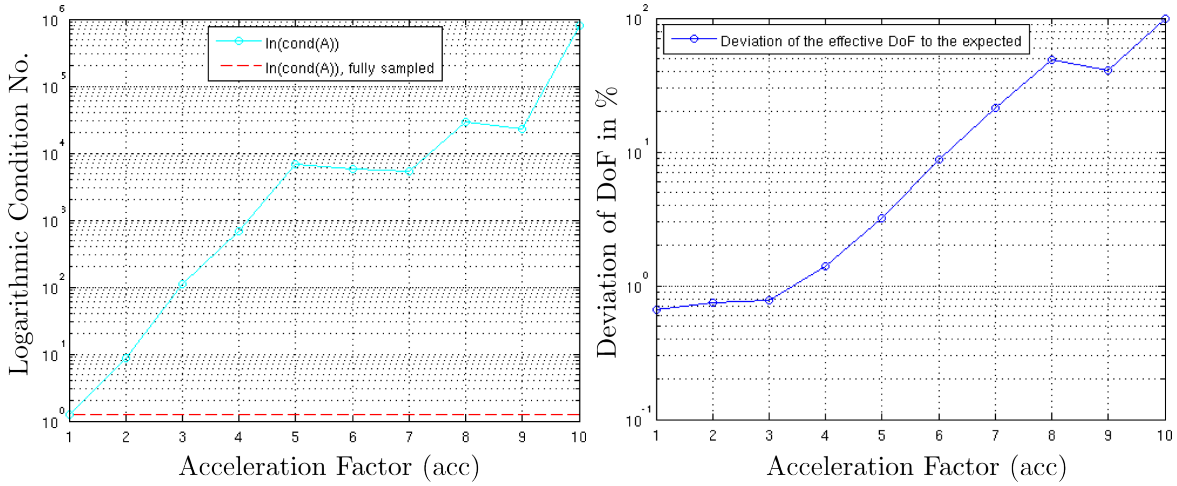


Figure 23: Left: Condition number of the problem: For the fully sampled case the condition number is 1.2338, then it increases approximately exponentially with the acceleration factor. Right: Deviation of the effective number of DoF with respect to the theoretically expected one. The kink at $acc = 3$ reflects the deteriorating visual impression of the images for $acc > 2$.

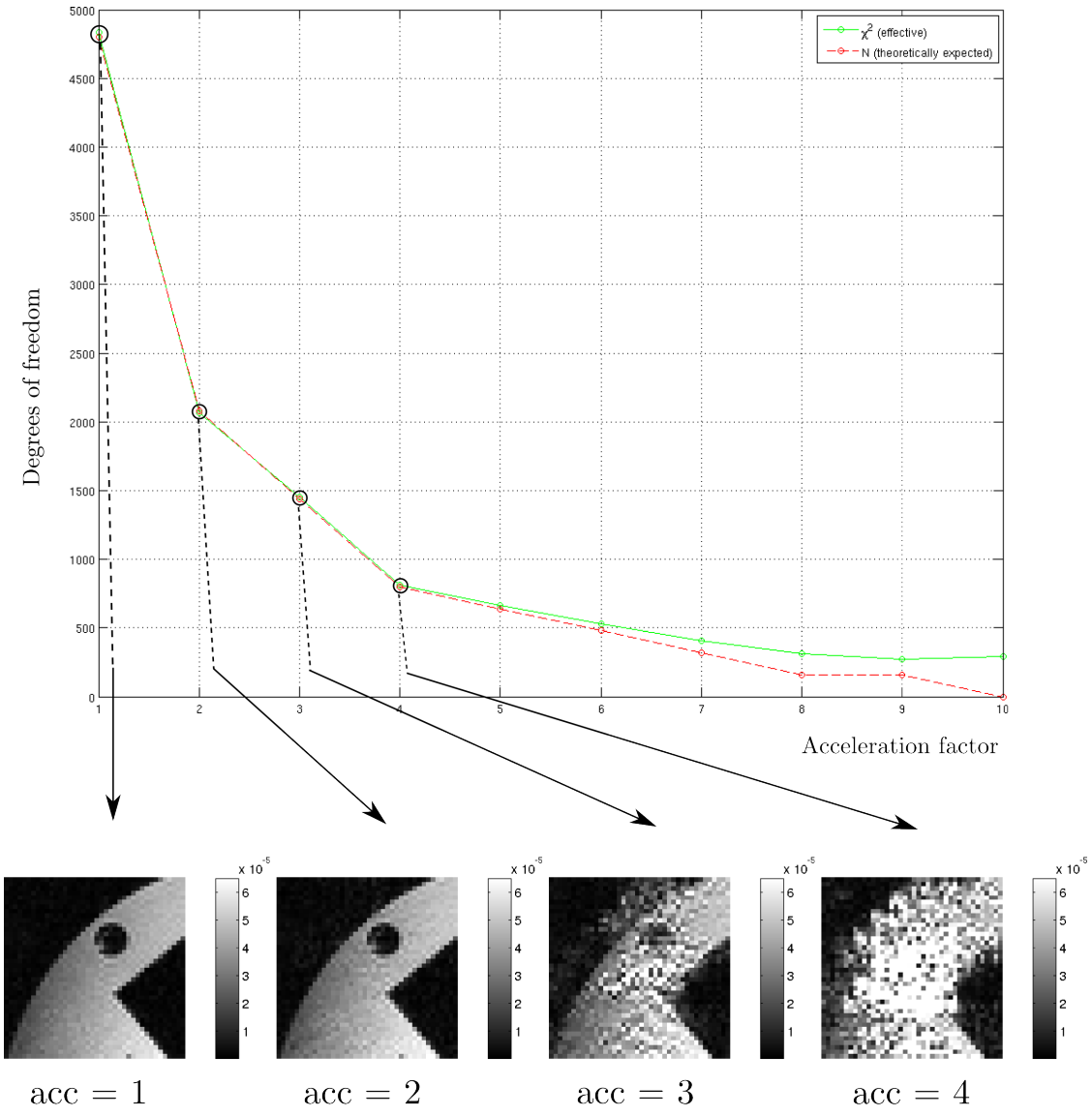


Figure 24: Model validation via misfit analysis for a dataset with $n_x = 40$, $n_y = 40$ measured by 4 coils: The misfit χ^2 (full line) is compared with the theoretically expected number of degrees of freedom N (dashed line) for different acceleration factors without regularization. For a given Gaussian noise statistics in k-space the reconstruction is shown (weighted CGSENSE). The effective number of DoF tends to be a bit higher than the expected one due to not fully linear independent coils.

5.3 Finding the optimal reconstruction parameter

The PDF for the regularization parameter of the Tikhonov approach has been derived in chapter 4.1 and has the shape

$$L(\alpha) := ln(p(\alpha|\mathbf{y}, C, \mathcal{I})) = C_0 + \frac{N-2}{2} ln(\alpha) + \frac{1}{2}\mathbf{y}^\dagger C^{-1}\mathcal{A}H^{-1}\mathcal{A}^\dagger C^{-1}\mathbf{y} - \frac{1}{2}\underbrace{ln(det(H))}_{tr(ln(H))}$$

Then the optimal parameter α_{opt} obeys the equation

$$\frac{2-N}{\alpha} = \frac{\partial}{\partial\alpha}\mathbf{y}^\dagger C^{-1}\mathcal{A}H^{-1}\mathcal{A}^\dagger C^{-1}\mathbf{y} - \frac{\partial}{\partial\alpha}tr(ln(H))$$

and can be determined directly if $H(\alpha)$ and $H^{-1}(\alpha)$ can be implemented efficiently as functions of α. To prove this principle a 100 x 100 dataset from 4 coils with acceleration factor 5 was taken. As we have seen in section 5.2 the problem is too ill conditioned so that the plain reconstruction leads to misleading images. That is why it needs to be regularized (here with the Tikhonov approach). A brute force method to obtain the optimal regularization parameter is to scan the parameter α in a large range and evaluate $L(\alpha)$. That avoids the generation of H and H^{-1} for every α and helps to get an idea of the behavior of the PDF. Figure 25 shows the region of the optimal α. The PDF for α) is similar to a Gaussian. The width of the peak region of $\sigma_\alpha \approx 0.2$ is very small compared to its location on the α-axis ($\alpha_{opt} = 9.62\ 10^8$).

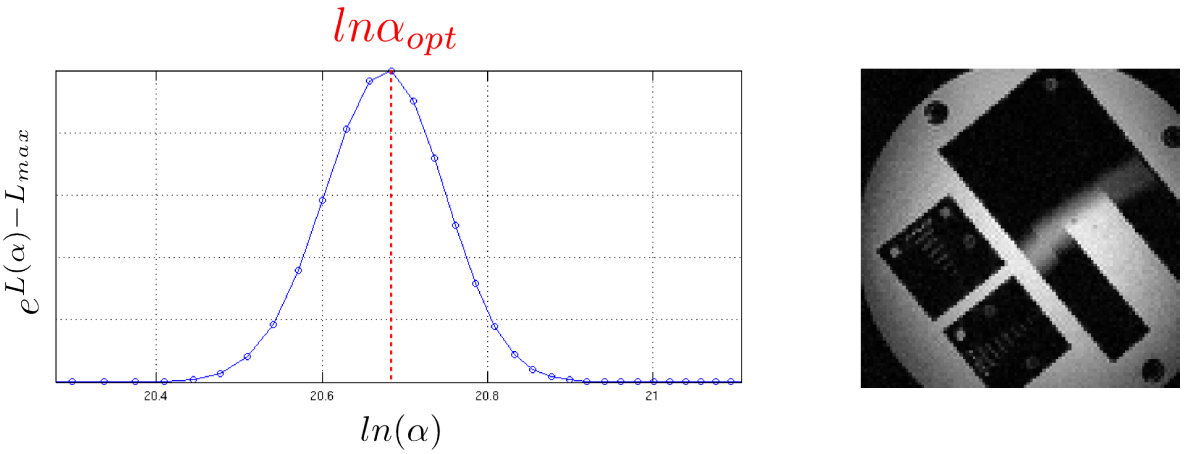


Figure 25: Detail of the peak of the log-PDF $L(\alpha)$ (left) with the reconstructed image for α_{opt} (right). Note the logarithmic scale!

The optimal reconstruction plotted in figure 25 looks pretty unblurred and without any visible noise or artifacts. However, there is no information about how the visual image quality gets when we move away from the optimum. That is why we zoom out a bit and plot the whole α-range in figure 26. The log-PDF $L(\alpha)$ is flat except in the small peak region which gives the hint that the solution for the optimal parameter is unique. The misfit χ^2 is also plotted and shows the transition between over- and underfitting. For most problems the first guess of the optimal parameter would be the turning point of χ^2. Here this is not the case: The optimal parameter is located in the region where χ^2 begins to grow. Looking at the reconstructed images, it turns out that for weak regularizations the image is governed by noise and aliasing artifacts in the vertical direction. The acceleration factor is 5 so we see 5 backfoldings. Shortly

before the region of α_{opt} is reached, the noise gets weaker and the structure of the measured phantom becomes visible. When approaching the vicinity of the peak region the visual image quality gets better and the aliasing effects disappear. As soon as α exceeds this region, the images get blurrier and other artifacts become visible. For $\alpha > \alpha_{opt}$ the structure of the phantom still remains but the numerical values for the pixels converge to zero (enhanced smallness). Here the Tikhonov term only allows solutions close to zero. One remarkable feature in this region is that the structure of the image is still preserved up to a certain level (not visible in figure 26 due to scaling), but the image can be more than 5 orders of magnitude smaller than in the optimal case.

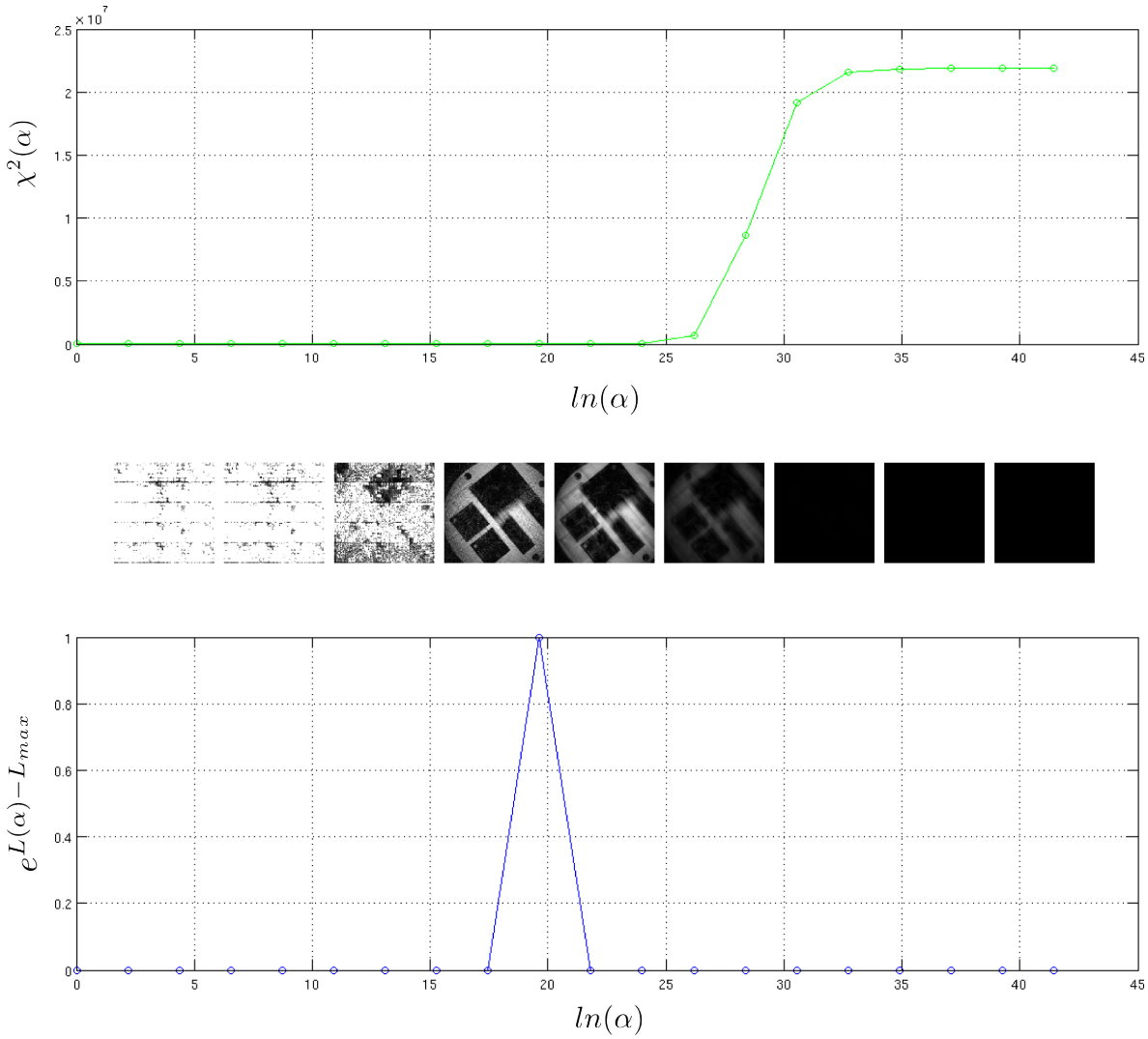


Figure 26: Overview of the misfit $\chi^2(\alpha)$ and the log-PDF $L(\alpha)$ for Tikhonov regularization with according image reconstructions. Note the logarithmic scale for the regularization parameter α. For all reconstructions the same scaling of the colormap has been used.

5.4 Evaluation of statistical quality measures

For the quality measures presented in section 4.4 either the PR-method or MCMC-sampling can be used to generate samples from which a covariance matrix $\mathrm{Cov}(\mathbf{x})$ can be estimated. In the following discussions the Tikhonov prior is used for the SENSE model to reconstruct a 128×128 dataset from $n_c = 4$ coils for acceleration factors acc = 1, 3 and 5. The reason for that is that for the Tikhonov approach (or more generally the signal power prior) mean and covariance can be computed analytically. However, a straight forward matrix inversion is needed for evaluation of the analytic expressions (see 4.10 for the PR-method and 4.12 for MCMC), which takes a lot of time if it is not optimized. That is why the following results have been sampled starting from a 128 x 128 x 4 dataset and not been computed via the analytic expression. Bayesian probability theory yields the correct posterior for this purpose but direct sampling via MCMC requires fine tuning and convergence control to obtain reliable results. That is why the PR-method is the sampling method of choice here even though it samples from a different posterior. Fortunately the means of both sampling methods are identical, only the covariance of the PR-method is underestimated compared to its true value found by Bayesian inference. Finally the reconstruction of each pseudo replica is done by minimizing the Tikhonov functional with the conjugate gradient method.

5.4.1 Noise distribution in complex images

For any reconstruction method the covariance matrix of magnitude images can be computed from the covariance matrix of the real/imaginary part of the complex images with the Rician distribution presented in equation 4.15 - as long as $\mathrm{Cov}(\mathbf{a}) \approx \mathrm{Cov}(\mathbf{b})$. If this is the case, also $\mathrm{Cov}(\mathbf{a}) \approx \mathrm{Cov}(\mathbf{b}) \approx 1/2\mathrm{Cov}(\mathbf{x})$ holds (see equation 5.1). Figure 27 shows $\mathrm{Cov}(\mathbf{a})$ and $\mathrm{Cov}(\mathbf{b})$ for acc = 1, 3 and 5: It is clear that the needed identity is given.

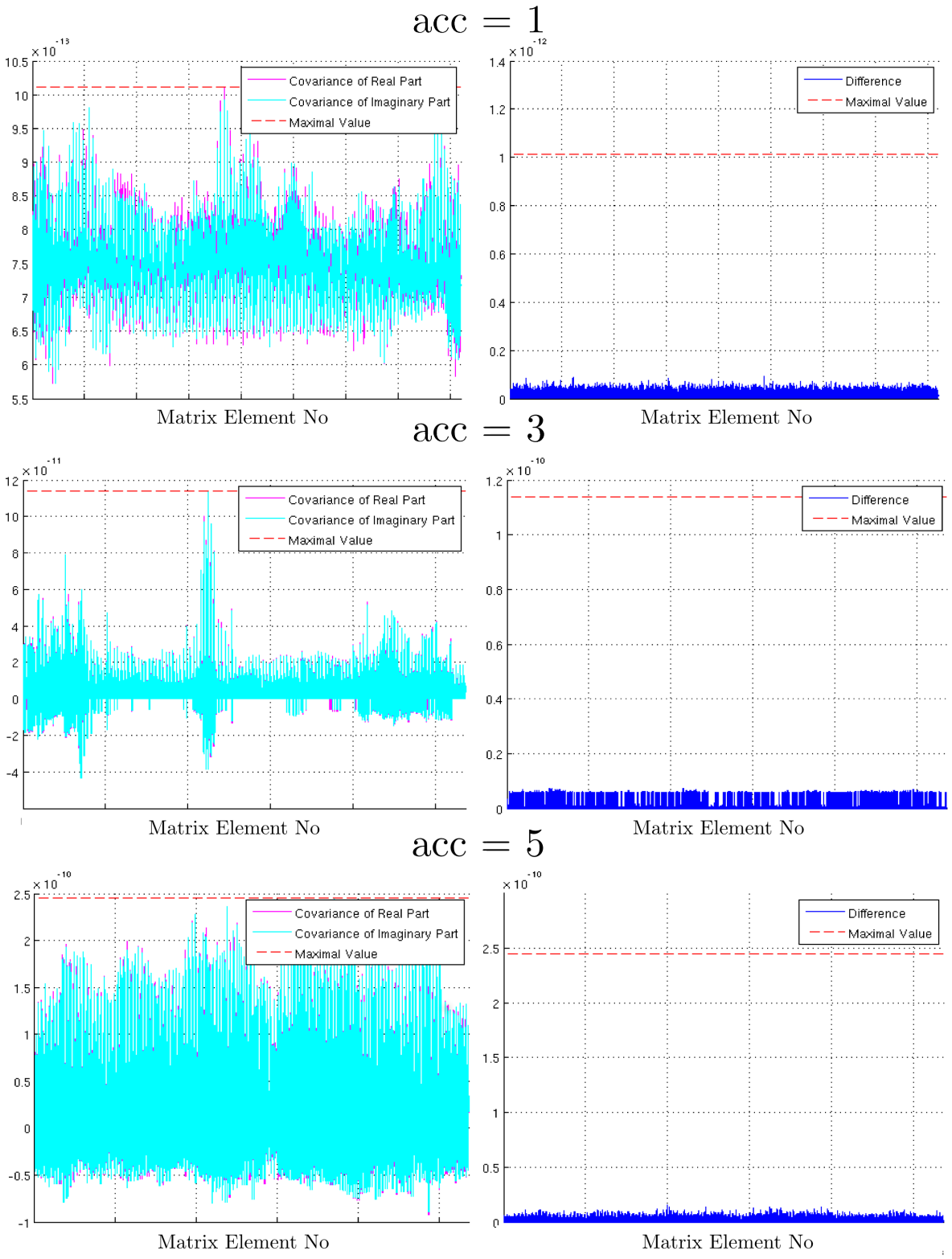


Figure 27: Left: The values of Cov(**a**) and Cov(**b**) are shown. Right: The difference between Cov(**a**) and Cov(**b**) is plotted and compared with the maximal value (red dashed line). It is below 10 % for all elements. That means that the covariance matrices of real and imaginary part of the image for acc = 1, 3, 5 are identical. Note that for acc = 1 only diagonal elements are plotted since they are the only non-zero ones.

A more detailed insight is the plot of the histograms of real and imaginary part of a single pixel value. Both show a Gaussian distribution with identical variance for each acceleration. This is another condition for the application of the Rician distribution. As one might expect the STD increases with the acceleration factor.

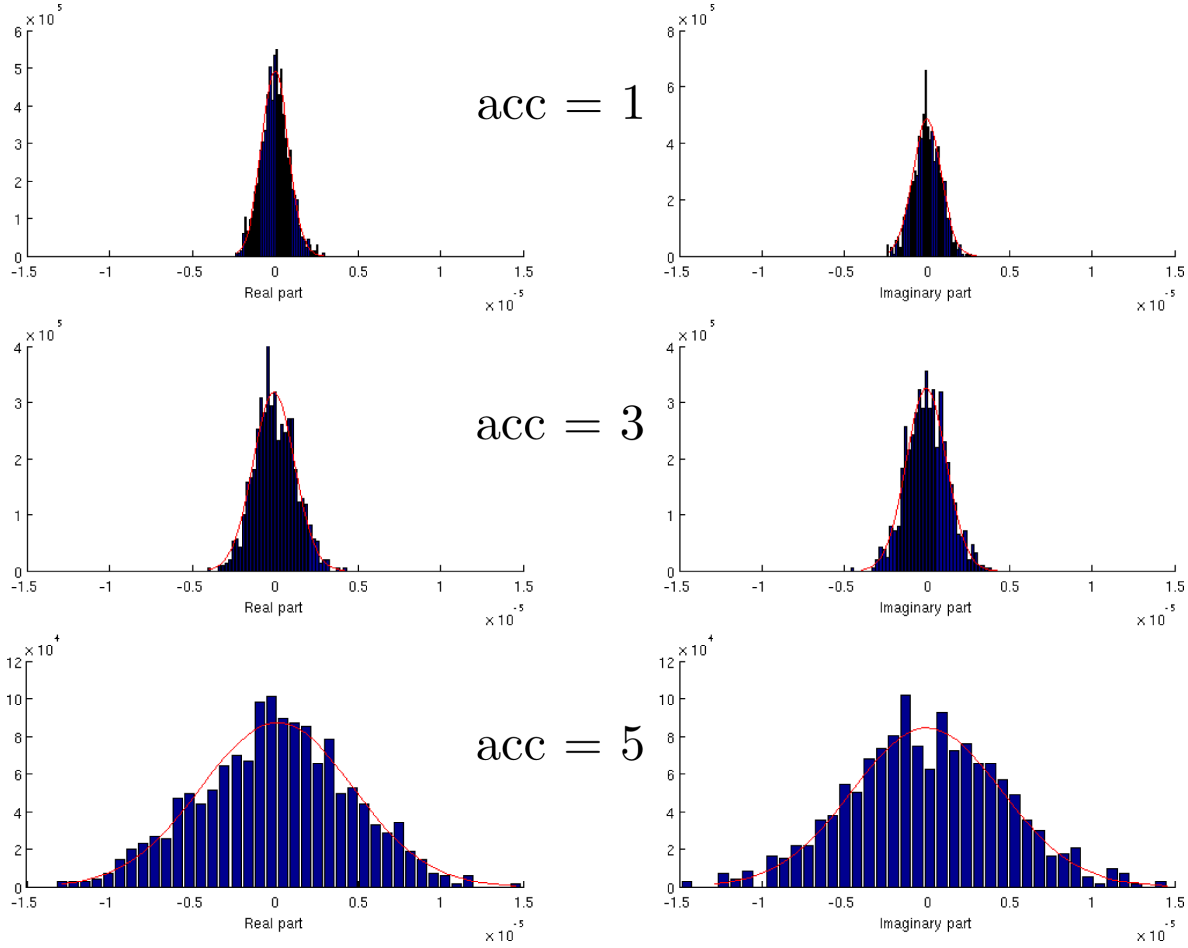


Figure 28: The distribution of a random pixel in real and imaginary part is shown for acc = 1, 3 and 5. Plotted is a not-normalized histogram for $x_i - \langle x_i \rangle$.

As a summary the following can be stated:

1. The calculation of the covariance of the magnitude with the Rician statistics according to 4.15 can be applied for the Tikhonov regularized SENSE model because the distribution of real and imaginary part of the pixels is Gaussian with identical covariance matrix.

Finally the covariance matrix for magnitude images can be computed. It turns out that the nonzero elements of the covariance matrices of real and imaginary part are sparsely centered around the main diagonal. For $acc = 1$ only the main diagonal is nonzero, for higher accelerations the nonzero elements are located on the main diagonal and s sub-diagonals where s is equal to $n_x = n_y$, the number of pixels in x or y direction of the image. A more detailed look identifies $N = n_x n_y$ square sub-blocks of the size $n_x \times n_y$. The following plots affirm that statement: Only elements greater than 0.1 times the maximal absolute value are considered which are sparsely concentrated around the main diagonal.

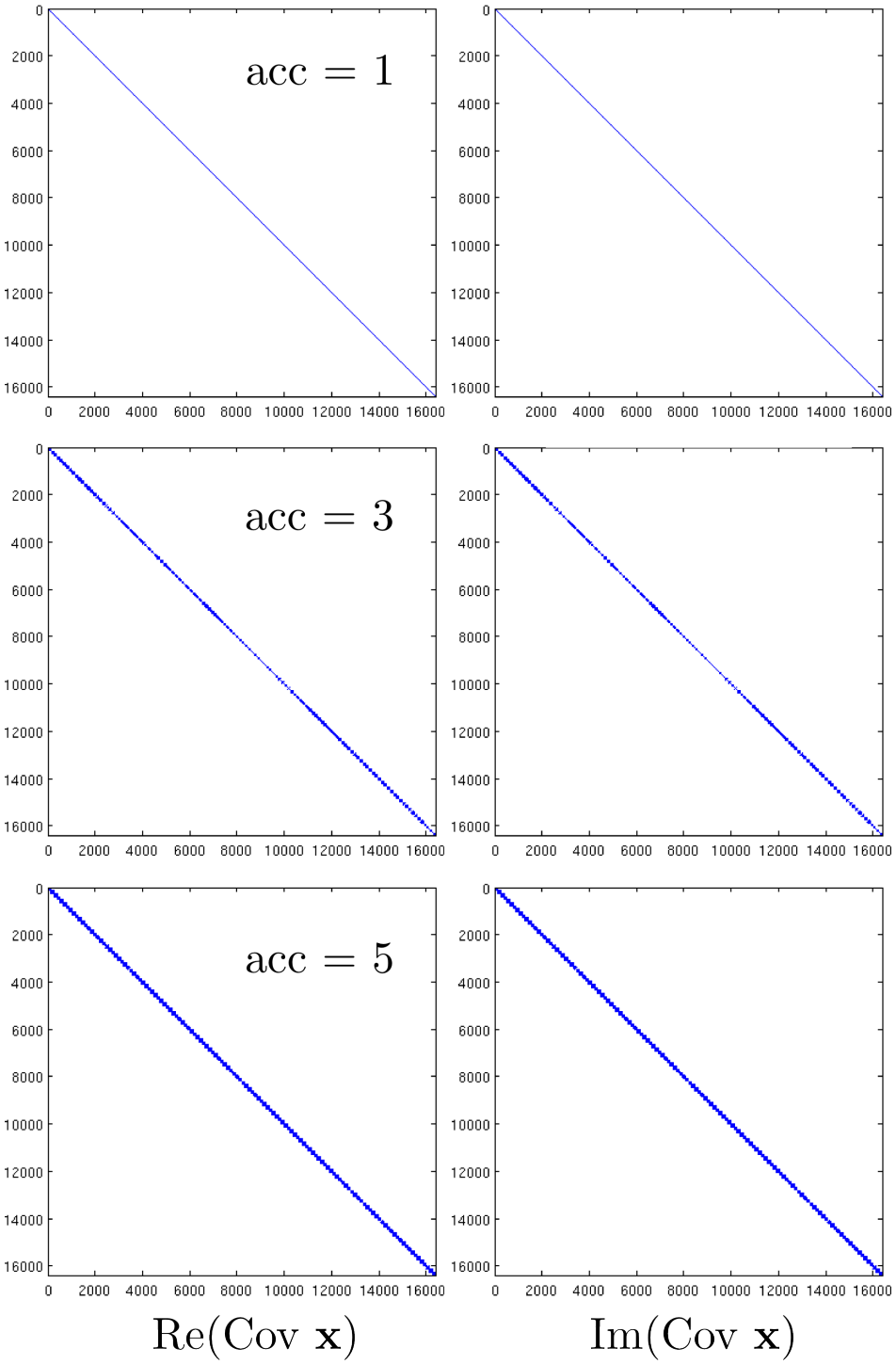


Figure 29: Structure of the covariance matrices of real and imaginary parts for different accelerations: The covariance matrices of real and imaginary part of the image for acc = 1, 3 and 5 are sparsely concentrated around the main diagonal.

That leads to another important result:

- The covariance matrix is sparsely centered around the main diagonal in $n_x \times n_y$ sub-blocks. That speeds up the calculation of the magnitude covariance by evaluating the Rician distribution.

In figure 30 the magnitude of the mean of the reconstructed 5000 pseudo replicas is shown for different accelerations. Next to it the STD maps according to the magnitude images are plotted to give an idea about pixelwise uncertainties even though this might be misleading because it does not take correlations into account (for more details see section 5.4.3). STD maps are somehow unpractical when it comes to the question of assessing the image quality in a certain ROI because for that the quantity of interest is actually the ratio of the pixels and their STDs. That is why in figure 31 the SNR maps of the magnitude for the same 3 datasets are shown. Here it is more obvious to see the quality: ROIs with high SNR have probably better quality than ROIs with low SNR.

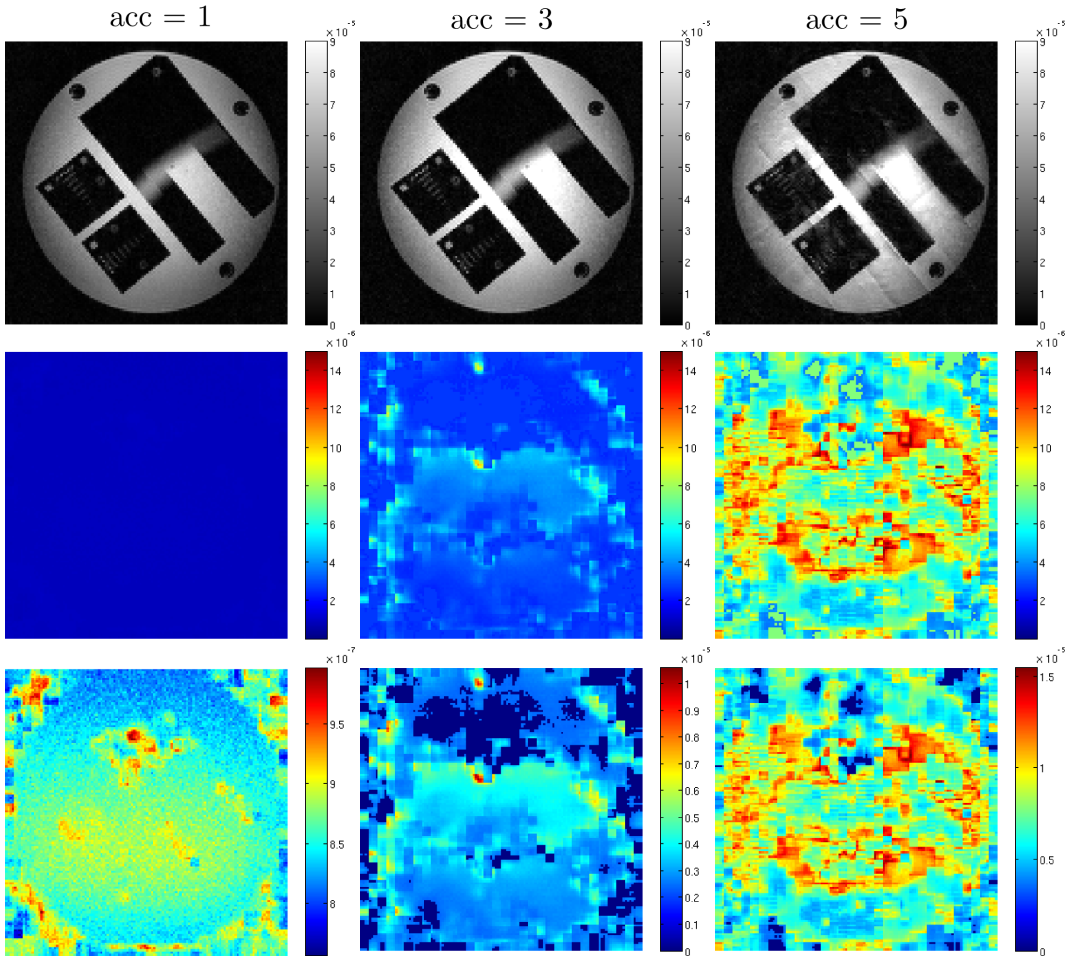


Figure 30: Rician STD maps of magnitude images for acceleration factors 1, 3 and 5 with the according means of 5000 reconstructed pseudo replicas: The first row shows the magnitude of the reconstruction, the second one the STD maps in the identical colormap and the third row shows the STD maps in their individually scaled colormaps for structural details.

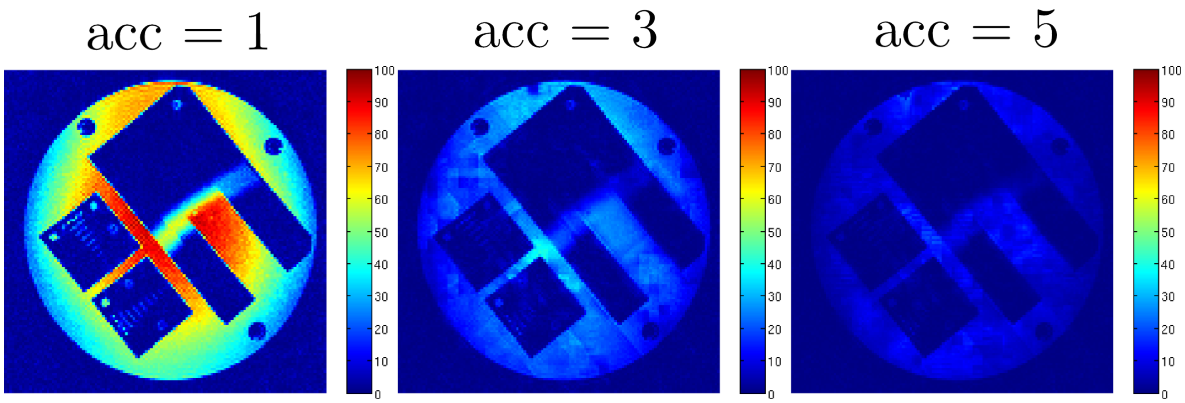


Figure 31: The pixelwise SNR-map of magnitude images (computed with the according Rician STD) is plotted for acceleration factor 1, 3 and 5 in the colormap of acc = 1, which has the highest SNR. With increasing acceleration the SNR gets lower. Regions of high SNR have tendentious better quality than low SNR regions.

In figure 32 the g-factor defined in equation 4.16 - which uses the not accelerated case as a gold standard - is also plotted for acc = 3 and acc = 5. It shows the map of noise amplification with respect to the unaccelerated case. This quality map is very useful - given the fact that a full reconstruction is available. In this case the quality of every pixel with values greater than 1 is worse than the optimal quality.

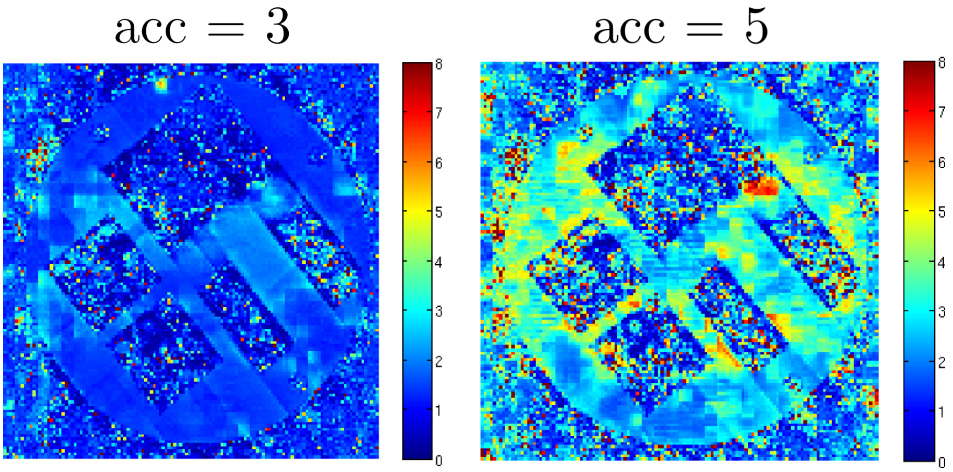


Figure 32: The pixelwise g-factor map of magnitude images (computed with the according Rician STD) is plotted for acceleration factor 3 and 5. It is basically the noise amplification factor compared to the reconstruction of the unaccelerated data set.

5.4.2 Comparison of Pseudo replica and MCMC

Based on Bayesian inference MCMC sampling is the method that correctly estimates the covariance and the mean for a certain dataset. An analysis of the image samples generated by MCMC turns out that they have again identical covariance in real and imaginary part of the image $\mathbf{x}$. A typical time evolution of a single pixel is shown in figure 33. The results of MCMC applied to the Tikhonov regularization with the SENSE model can be directly

compared with the PR method according to equations 4.10 and 4.12. Both methods return the same mean but their covariances differ: The PR method's covariance is always smaller or equal to the MCMC's because $H^{-1}BH^{-1} \geq 0$. The reason for that is because the PR-method yields only an approximation for the true posterior of the problem. The following equation summarizes the results again:

$$\mathrm{Cov}_{PR}(\mathbf{a}) \approx \mathrm{Cov}(\mathbf{b}) \approx \frac{1}{2}\mathrm{Cov}_{PR}(\mathbf{x}) = \frac{1}{2}\left(H^{-1} - H^{-1}BH^{-1}\right)$$
$$\mathrm{Cov}_{MCMC}(\mathbf{a}) \approx \mathrm{Cov}(\mathbf{b}) \approx \frac{1}{2}\mathrm{Cov}_{MCMC}(\mathbf{x}) = \frac{1}{2}H^{-1}$$

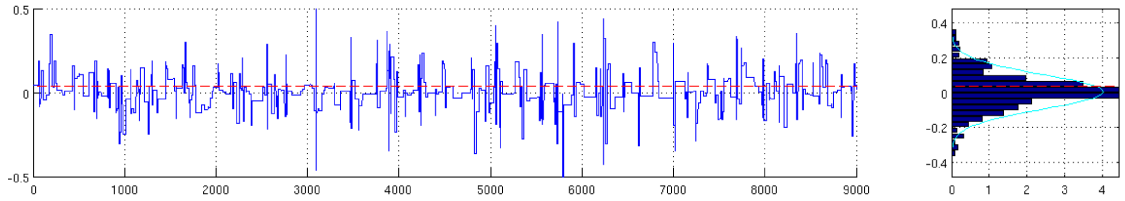


Figure 33: Time evolution of the real part of a single pixel during the evaluation phase of MCMC: The pixel (blue line) approaches a stationary distribution (cyan line) whose mean converges to the true ground truth value (red dashes) which was used to verify the results of the MCMC algorithm during coding.

In figure 35 the averages of both the PR method and the MCMC method are compared. The PR-method is easier to apply in practice because it does not require any fine tuning, Burn-in selection, optimal scaling of the step size and convergence control like MCMC. PR-sampling does not yield the correct covariance matrix but the the result for the mean is correct. The mean of the MCMC run were obtained from 500 samples, whereas the PR method's mean was extracted from 5000 samples. That is due to the fact that only every $n_x n_y$ steps every pixel is changed (or at least got a new proposal state) which makes the MCMC run unfortunately very time consuming. The MCMC reconstructions are very accurate: Even the phase image is reconstructed accurately. In figure 34 the STD maps for magnitude images of both methods are compared after the covariance of the real/imaginary part has been transformed with the Rician distribution.

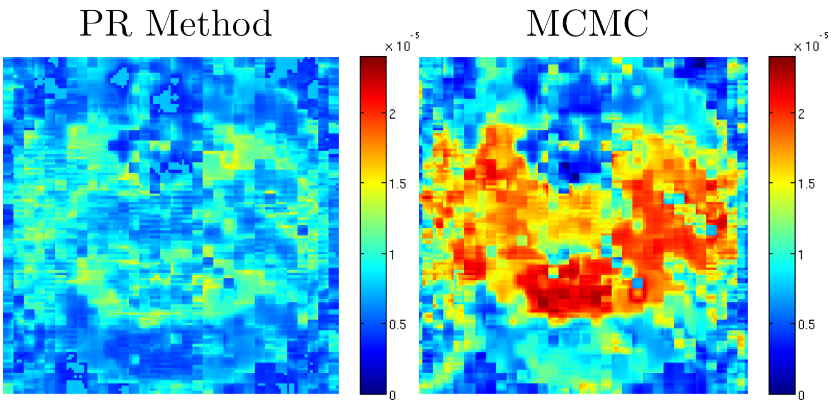


Figure 34: Comparison of the STD maps from PR method and MCMC: The PR method results in lower STD than MCMC, which is generally expected by the analytic results in equation 5.2.

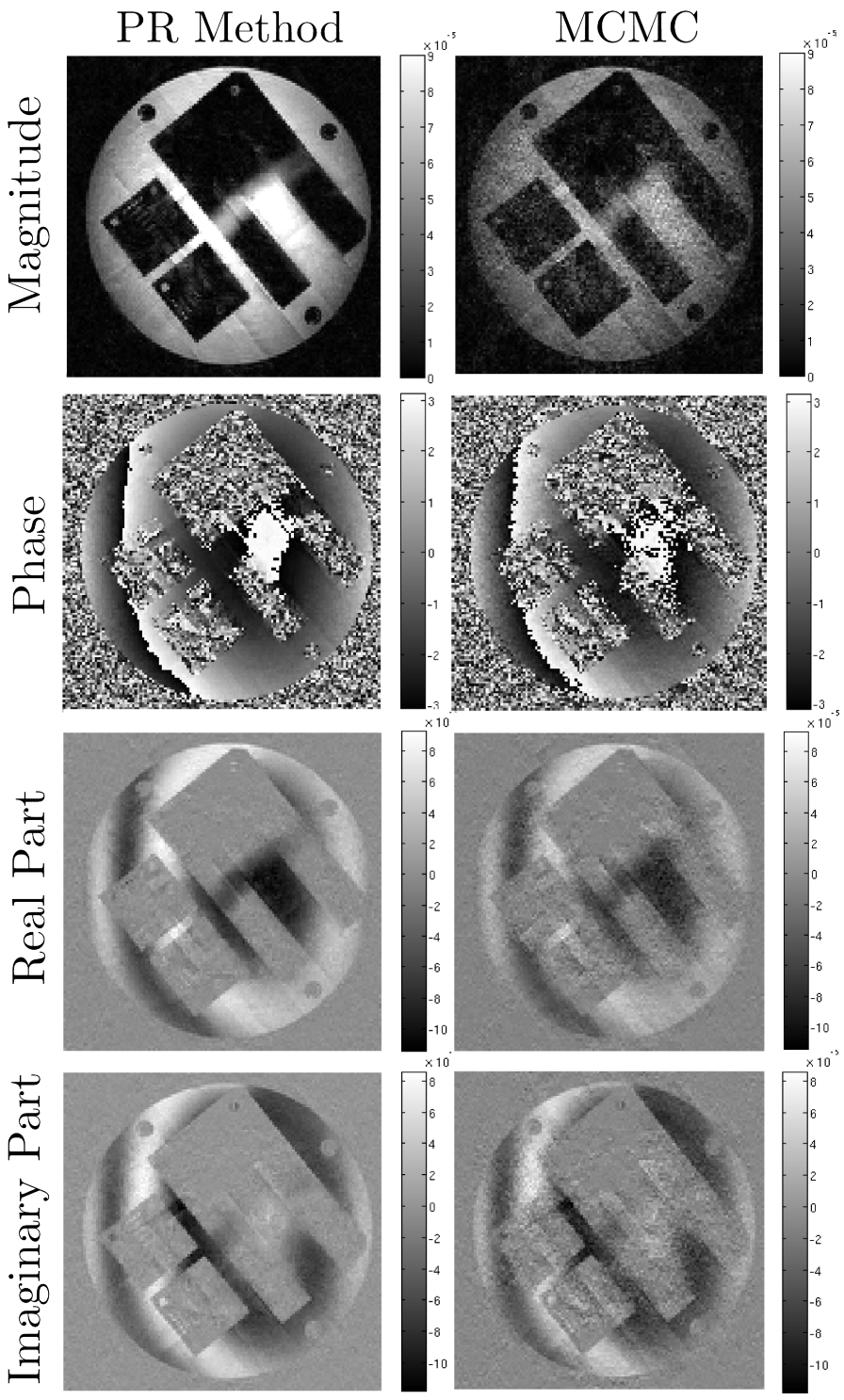


Figure 35: Comparison of reconstructions from PR method and MCMC for a 128 x 128 dataset from 4 coils at acceleration factor 5: The left column shows the means of 5000 samples obtained by the PR-method. In the right column the means of 500 samples from a MCMC run of the same dataset is plotted. The MCMC reconstruction is still a bit noisy because compared to the PR-method only a 10^{th} of the samples were used and the chain did not converge fully.

5.4.3 Evaluation of the covariance quality measure

Now the quality measure based on the covariance of the magnitude image - which was presented in equation 4.17 - is evaluated on the full image and on 2 different ROIs. Again the same 128 x 128 datasets from 4 coils is used. The measure is both valid for correlated and uncorrelated pixels. As pointed out in figure 29, the covariance matrices are not diagonal for accelerations greater than 1. Therefore it is expected that for accelerated datasets the evaluation of the quality measure using the full covariance matrix will differ from the case when only its main diagonal (the variances) are used. This consideration is confirmed. In figure 32 the quality for reconstructions of datasets acquired at acceleration factor 1, 3 and 5 is shown. For each of these three datasets the quality obtained by the variance and the covariance measure is compared for 3 different ROIs: the full image, Region 1 and Region 2 (see figure 36).

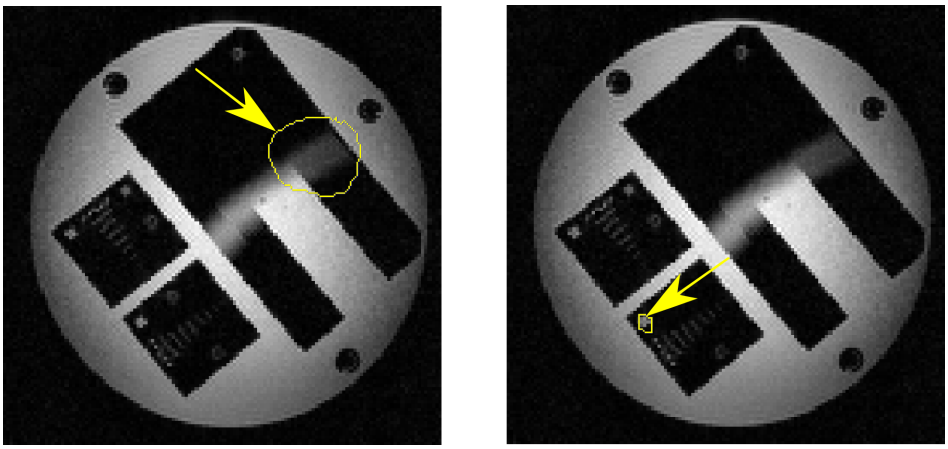


Figure 36: Two different ROIs to test the quality measures in which the covariance and the variance quality measure was evaluated: Region 1 (left) and Region 2 (right).

It turns out that the quality decreases generally for increased acceleration as one might expect. On top of that it underlines the fact that pixel correlations are important to take into account when assessing the quality of an image of higher accelerations than $acc = 1$. If the variance based quality of the images - which neglects pixel correlations - is compared with the "true" one obtained by the covariance in the same ROI, it turns out that the variance based quality is always higher. That is because pixel correlations come into play and lower the quality. That can be misleading if one wants to assess the quality of a ROI based on the variance only. The comparison between the case of using the diagonal elements of the covariance only and

acc	Full		Region 1		Region 2	
-	Cov	Var	Cov	Var	Cov	Var
1	2.8469e-6	2.8469e-6	0.12394e-3	0.12394e-3	2.3e-3	2.3e-3
3	1e-3	7.0091e-6	7.4e-3	0.29769e-3	28.9e-3	6.5e-3
5	2.6e-3	17.654e-6	20.4e-3	1.5e-3	3e-3	1e-3

Table 1: Quality according to equation 4.17 for the case of neglected correlations (Var) and the one with included (Cov) for different acceleration factors in the full image and 2 other ROIs.

the case of neglecting them shows significant differences for datasets with acceleration factor

higher than 1.

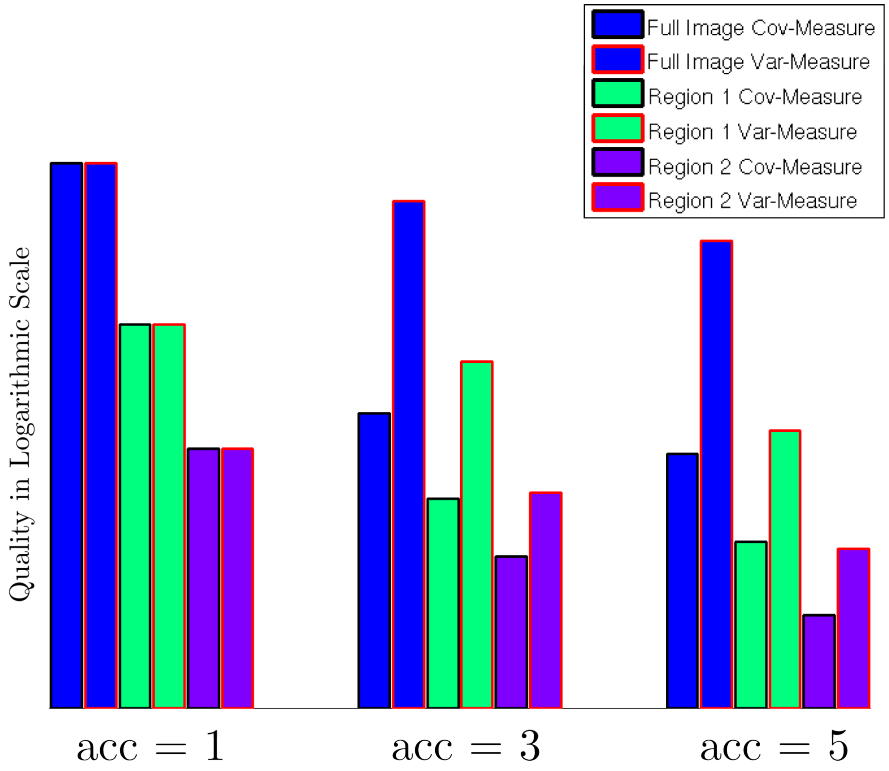


Figure 37: Quality of reconstruction for variance and covariance measure in different ROIs for acceleration factor 1, 3 and 5: The quality obtained by the variance measure is always higher than the quality measured by the covariance measure because the first ignores pixel correlations.

5.5 Summary of the results and conclusion

- **Optimal regularization parameter for the signal power prior**
  From the Bayesian point of view the optimal regularization parameter has been derived for the SENSE model of parallel imaging for the signal power prior approach. The general solution was applied to Tikhonov regularization but can also be used for priors using the derivative or curvature of the image by simply replacing a matrix. The optimal parameter depends on the k-space data, the covariance of k-space noise and the coil sensitivities and can be used for any acceleration factor. Given those input variables in 4.5 an equation is presented which can be solved for the optimal regularization parameter. In the field of MR image reconstruction this is a new result.

- **Analysis of noise samples in k-space and image space and sampling methods to obtain image samples**
  The known result that k-space noise is uncorrelated Gaussian noise with identical covariance matrices of real and imaginary part of the signal has been shown. Two different known methods to generate image samples from a single k-space dataset are presented and used: Markov Chain Monte Carlo sampling using the posterior predicted

by Bayesian inference and the bootstrap approach via Pseudo Replica sampling. For both methods the posterior probability distributions as well as their means and covariances are derived analytically (see equations 4.10 and 4.12). It turns out that the image samples are also multivariate Gaussians and that the means of the images generated from the MCMC and the PR sampling are identical but the covariance is underestimated for the PR-method. The reason for the error of the PR-method is that it does not exactly sample from the "true" posterior predicted by Bayesian inference. In general the covariance matrix in image space is not uncorrelated anymore. Only for the fully sampled k-space this is still the case. That means that in image space pixel correlation effects play a role and cannot be neglected for accurate results.

- **Derivation of the Rician distribution for the magnitude of Gaussian random variables**
  Since the magnitude of the complex valued image is the representation that is mostly used in clinical practice, the covariance matrices calculated for the images samples - which consist of complex valued pixels - need to be converted to the covariance of the magnitude images. This is achieved for single pixels by using the known Rician distribution. However, for correlated pixels this approach does not work anymore. That is why a general derivation of the PDF for magnitudes obtained from correlated complex images is given in equation 4.15 - a new result in MRI.

- **Application of the covariance quality measure for MR reconstructions and pixel correlations**
  To save the time needed to measure and reconstruct multiple datasets to get information about its statistical behavior the PR and the MCMC sampling methods are able to produce samples in image space from only one single dataset. From these samples the covariance matrix can be estimated and the covariance measure presented in equation 4.17 can be used to asses the relative reconstruction quality of an image. Compared to the widely used STD or SNR maps it has the additional advantage that it takes pixel correlations into account. Only in the unaccelerated case the STD or SNR map already contains the full information about uncertainties in the image. For accelerated datasets the use of variances only can lead to misleading inferences and their usage has to be considered and justified thoroughly. This fact was underlined with an example where the quality of a certain ROI was evaluated by taking the off-diagonal elements of the covariance into account and by neglecting them. The comparison confirms the concern.

- **General considerations about quality measures**
  The wish of assessing the quality of a reconstructed image is obvious given various reasons presented in the introduction. For the sake of generality a quality measure based on a gold standard needs to be avoided because it biases the outcome with the chosen standard. However, in MRI there still exists an unbiased gold standard which is the unaccelerated image. For a fully sampled dataset - given the fact that the SNR is sufficiently high to avoid further ill-conditioning - the image can be reconstructed with almost absolute reliability because the MRI model is known and the experimental implementation is sophisticated. That is the reason why unaccelerated MRI techniques are licensed for clinical use. Therefore the reconstruction from the according fully sampled dataset is the perfect gold standard. But for assessing the quality of MRI sequences

with increased speed due to undersampling it is still useless because a fully sampled k-space is always always required. That leads to the idea of developing a relative quality measure. In the statistical sense such is given by the covariance quality measure. Its additional advantage next to its applicability for undersampling protocols is that it takes pixel correlation into account - which is not the case in g-factor maps. It returns a scalar value for a given ROI that basically compares the cumulative amount of uncertainty with the cumulative amount of signal and can therefore be used to differentiate between "real" and "noisy" structures in the image. Unfortunately it does not provide any absolute reference. Even if the optimal regularization parameter for a certain combination of regularization, data, undersampling and noise has been chosen and also the quality of a certain ROI is good according to the covariance quality measure, the image can still have undesirable artifacts that are not "real". The reasons for that is the the wrong image prior or regularization is used that either can not compensate the ill-conditioning of the problem anymore or introduces artifacts into the solutino itself. At this point a conceptual problem comes up: If we knew the perfect prior that would not have any of the described properties, we would already have a lot of information about the solution. In other words, there exists some kind of conservation law of information: If we do not have enough information about the problem (little data, ill conditioned problem matrix) we need to balance this lack of information by choosing a "good" prior. The same problem occurs with quality measures: In the extreme case of knowing the solution we have the perfect measure in our hands - but it is useless. If we do not know the solution but we want to assess the quality of a guess-solution, we do not have a proper quality measure. That means: Having a reliable quality measure for guess-solutions is equivalent to the optimal solution itself. As a conclusion it can be stated that the presented statistical quality measures are only useful within the frame that is spanned by the problem formulation itself - which is MRI model, prior and noise statistics - and that they cannot go beyond that.

6 Appendix

6.1 Efficient evaluation of the autocorrelation function

This section follows the lecture notes *Computer Simulations* [6]. The autocorrelation function for M-dimensional random variables is defined as

$$\rho(t) := \frac{Cov(\mathbf{X}_j, \mathbf{X}_{j+t})}{\sigma_j \sigma_{j+t}}$$

Rewritten for finite sample sizes we obtain the empirical autocorrelation function as

$$\begin{aligned}\rho^E(t) &:= \frac{\sum_{j=1}^{N-t} \left(\mathbf{X}_j - \overline{\mathbf{X}}(t)\right)^T \left(\mathbf{Y}_j - \overline{\mathbf{Y}}(t)\right)}{\sqrt{\left(\sum_{j=1}^{N-t} \left(\mathbf{X}_j - \overline{\mathbf{X}}(t)\right)^2\right)\left(\sum_{j=1}^{N-t} \left(\mathbf{Y}_j - \overline{\mathbf{Y}}(t)\right)^2\right)}} \\ &= \frac{\sum_{j=1}^{N-t} \mathbf{X}_j^T \mathbf{Y}_j - \mathbf{X}_j^T \overline{\mathbf{Y}}(t) - \overline{\mathbf{X}}^T(t) \mathbf{Y}_j + \overline{\mathbf{X}}^T(t) \overline{\mathbf{Y}}(t)}{\sqrt{\left(\sum_{j=1}^{N-t} \mathbf{X}_j^2 - 2\mathbf{X}_j^T \overline{\mathbf{X}}(t) + \overline{\mathbf{X}}^2(t)\right)\left(\sum_{j=t+1}^{N} \mathbf{X}_j^2 - 2\mathbf{X}_j^T \overline{\mathbf{X}}(t) + \overline{\mathbf{X}}^2(t)\right)}}\end{aligned}$$

where we will define the averages as

$$\begin{aligned}\overline{\mathbf{X}}(t) &= \frac{1}{N-t} \sum_{j=1}^{N-t} \mathbf{X}_j \\ \overline{\mathbf{Y}}(t) &= \frac{1}{N-t} \sum_{j=1}^{N-t} \mathbf{Y}_j = \frac{1}{N-t} \sum_{j=t+1}^{N} \mathbf{X}_j\end{aligned}$$

This formulation is called "two-pass algorithm" in literature. Now we replace all the terms regarding the random variable $\mathbf{Y}_j$ with the according terms in $\mathbf{X}_j$ and plug in definitions of

the averages. We get for the denominator:

$$\begin{aligned}
\sum_{j=1}^{N-t} \mathbf{X}_j^T \overline{\mathbf{X}}(t) &= \sum_{j=1}^{N-t} \mathbf{X}_j^T \frac{1}{N-t} \sum_{j=t+1}^{N} \mathbf{X}_j \\
&= \frac{1}{N-t} \sum_{j=1}^{N-t} \mathbf{X}_j^T \sum_{j=t+1}^{N} \mathbf{X}_j \\
\sum_{j=1}^{N-t} \overline{\mathbf{X}}^T(t) \mathbf{X}_{j+t} &= \frac{1}{N-t} \sum_{j=1}^{N-t} \mathbf{X}_j^T \sum_{j=t+1}^{N} \mathbf{X}_j \\
\sum_{j=1}^{N-t} \overline{\mathbf{X}}^T(t) \overline{\mathbf{X}}(t) &= (N-t) \overline{\mathbf{X}}^T(t) \overline{\mathbf{X}}(t) \\
&= \frac{N-t}{N-t} \sum_{j=1}^{N-t} \mathbf{X}_j^T \frac{1}{N-t} \sum_{j=t+1}^{N} \mathbf{X}_j \\
&= \sum_{j=1}^{N-t} \mathbf{X}_j^T \frac{1}{N-t} \sum_{j=t+1}^{N} \mathbf{X}_j \\
&= \frac{1}{N-t} \sum_{j=1}^{N-t} \mathbf{X}_j^T \sum_{j=t+1}^{N} \mathbf{X}_j
\end{aligned}$$

This shows that two terms cancel out. For the argument of the square-root of the nominator we get in similar fashion:

$$\begin{aligned}
&\left(\sum_{j=1}^{N-t} \mathbf{X}_j^2 - 2 \sum_{j=1}^{N-t} \mathbf{X}_j^T \overline{\mathbf{X}}(t) + \sum_{j=1}^{N-t} \overline{\mathbf{X}}(t)^2 \right) \left(\sum_{j=t+1}^{N} \mathbf{X}_j^2 - 2 \sum_{j=t+1}^{N} \mathbf{X}_j^T \overline{\mathbf{X}}(t) + \sum_{j=t+1}^{N} \overline{\mathbf{X}}(t)^2 \right) = \\
&\left(\sum_{j=1}^{N-t} \mathbf{X}_j^2 - 2 \frac{1}{N-t} \sum_{j=1}^{N-t} \mathbf{X}_j^T \sum_{j=1}^{N-t} \mathbf{X}_j + \frac{N-t}{N-t} \sum_{j=1}^{N-t} \mathbf{X}_j^T \frac{1}{N-t} \sum_{j=1}^{N-t} \mathbf{X}_j \right) \\
&\left(\sum_{j=t+1}^{N} \mathbf{X}_j^2 - 2 \frac{1}{N-t} \sum_{j=t+1}^{N} \mathbf{X}_j^T \sum_{j=t+1}^{N} \mathbf{X}_j + \frac{N-t}{N-t} \sum_{j=t+1}^{N} \mathbf{X}_j^T \frac{1}{N-t} \sum_{j=t+1}^{N} \mathbf{X}_j \right) = \\
&\left(\sum_{j=1}^{N-t} \mathbf{X}_j^2 - \frac{1}{N-t} \sum_{j=1}^{N-t} \mathbf{X}_j^T \sum_{j=1}^{N-t} \mathbf{X}_j \right) \left(\sum_{j=t+1}^{N} \mathbf{X}_j^2 - \frac{1}{N-t} \sum_{j=t+1}^{N} \mathbf{X}_j^T \sum_{j=t+1}^{N} \mathbf{X}_j \right)
\end{aligned}$$

Canceling unnecessary terms yields a simplified result for the autocorrelation function:

$$\rho^E(t) = \frac{\sum_{j=1}^{N-t} \mathbf{X}_j^T \mathbf{X}_{j+t} - \frac{1}{N-t} \sum_{j=1}^{N-t} \mathbf{X}_j^T \sum_{j=t+1}^{N} \mathbf{X}_j}{\sqrt{\left(\sum_{j=1}^{N-t} \mathbf{X}_j^2 - \frac{1}{N-t} \sum_{j=1}^{N-t} \mathbf{X}_j^T \sum_{j=1}^{N-t} \mathbf{X}_j \right) \left(\sum_{j=t+1}^{N} \mathbf{X}_j^2 - \frac{1}{N-t} \sum_{j=t+1}^{N} \mathbf{X}_j^T \sum_{j=t+1}^{N} \mathbf{X}_j \right)}}$$

With the following definitions these typographic monsters get a bit prettier:

$$\mathbf{a} := \sum_{j=1}^{N-t} \mathbf{X}_j$$

$$\mathbf{b} := \sum_{j=t+1}^{N} \mathbf{X}_j$$

$$\rho^E(t) = \frac{\sum_{j=1}^{N-t} \mathbf{X}_j^T \mathbf{X}_{j+t} - \frac{1}{N-t}\mathbf{a}^T\mathbf{b}}{\sqrt{\left(\sum_{j=1}^{N-t} \mathbf{X}_j^2 - \frac{1}{N-t}\mathbf{a}^T\mathbf{a}\right)\left(\sum_{j=t+1}^{N} \mathbf{X}_j^2 - \frac{1}{N-t}\mathbf{b}^T\mathbf{b}\right)}}$$

Since the sums in **a** and **b** overlap they can be computed conveniently with a cumulative-sum command. In the same way the sum of squares can be computed. Only the first sum over mixed terms in the denominator needs further attention since it carries computational costs of order N^2. If we rewrite $\mathbf{X}_j$ as a vector function in time $\mathbf{f}(t_j)$ the sum represents a convolution in the time domain.

$$\sum_{j=1}^{N-t} \mathbf{X}_j^T \mathbf{X}_{j+t} = \sum_{j=1}^{N-t} \mathbf{f}^T(t_j)\mathbf{f}(t_j + t)$$

After Fourier-transforming the sum into frequency space the convolution becomes a simple product. Then the inverse Fourier-transform is taken to get the sum we were looking for. This trick brings the costs for calculating the sum down to the order of maximal $2Nlog(N)$ for FFT/IFFT.

One further consideration has to be made: The convolution theorem works for periodic functions. Since the random variable $\mathbf{X}_j$ is not periodic, it has to be modified by zero-padding that it becomes periodic. We assume a periodicity with period L, where L is defined in the range $2N \geq L \geq N + t_{max}$. Now we can replace $\mathbf{f}(t_j)$ with the zero-padded function $\mathbf{F}(t_j)$ which is defined as

$$\mathbf{F}(t_j) = \begin{cases} \mathbf{f}(t_j) & 1 \leq j \leq N \\ \mathbf{0} & N+1 \leq j \leq L \end{cases}$$

That makes the convolution theorem work in the desired range of arguments. **Note:** The absolute value in the equation below is calculated element-wise and is not the length of the

vector! We can write

$$\begin{aligned}\sum_{j=1}^{N-t}\mathbf{X}_j^T\mathbf{X}_{j+t} &= \sum_{j=1}^{N-t}\mathbf{f}^T(t_j)\mathbf{f}(t_j+t)\\ &= \sum_{j=1}^{L}\mathbf{F}^T(t_j)\mathbf{F}(t_j+t)\\ &= \sum_{j=1}^{L}\sum_{k=1}^{N}F_k(t_j)F_k(t_j+t)\\ &= \sum_{k=1}^{N}\sum_{j=1}^{L}F_k(t_j)F_k(t_j+t)\\ &= \sum_{k=1}^{N}\sqrt{L}\frac{1}{\sqrt{L}}\sum_{j=1}^{L}e^{i\omega_{ij}t}\left|\tilde{f}_k(\omega_{ij})\right|^2\end{aligned}$$

This operation has computational costs of $Llog(L)$ which can be maximal $2Nlog(2N)$ since $L \leq 2N$.

6.2 Efficient evaluation of the acceptance rate

For the computation of the acceptance rate the posterior given in equation 4.11 has to be evaluated for the old and the newly proposed state. This can be simplified considerably by rearranging terms and precalculating its all possible values of its effective expression. This is shown for Likelihood and prior individually. In the following it is assumed that the inverse covariance matrix C^{-1} has already been included into data and model such that the resulting covariance is the identity matrix.

6.2.1 Likelihood (flat prior)

For posterior distribution consisting of the Likelihood-term only (with flat prior) the acceptance rate becomes:

$$Acc = \frac{p(\mathbf{x}_n)q(\mathbf{x}_a|\mathbf{x}_n)}{p(\mathbf{x}_a)q(\mathbf{x}_n|\mathbf{x}_a)}$$

When the proposal distribution is chosen symmetric such that $q(\mathbf{x}_n|\mathbf{x}_a) = q(\mathbf{x}_a|\mathbf{x}_n)$, e.g. as a Gaussian or Uniform distribution, the proposal terms cancel out and the acceptance ratio gets

$$\begin{aligned}Acc &= \frac{e^{-\frac{1}{2}\left|\left|\mathbf{y}-\mathcal{A}\mathbf{x}_n\right|\right|_2^2}}{e^{-\frac{1}{2}\left|\left|\mathbf{y}-\mathcal{A}\mathbf{x_n}\right|\right|_2^2}}\\ &= e^{-\frac{1}{2}\left(\left|\left|\mathbf{y}-\mathcal{A}\mathbf{x}_n\right|\right|_2^2-\left|\left|\mathbf{y}-\mathcal{A}\mathbf{x}_a\right|\right|_2^2\right)}\\ &= e^{-\frac{1}{2}\Phi_0}\end{aligned}$$

If we write $\mathbf{x}_n$ as the sum of the old vector and a difference vector $\mathbf{x}_n = \mathbf{x}_a + \boldsymbol{\Delta}$ and define $\mathbf{v} := \mathbf{y} - \mathcal{A}\mathbf{x}_a$ and $\mathbf{z} := \mathcal{A}\boldsymbol{\Delta}$, Φ_0 can be rewritten as

$$\begin{aligned}
\Phi_0 &= \Big|\Big|\mathbf{y} - \mathcal{A}\mathbf{x}_n\Big|\Big|_2^2 - \Big|\Big|\mathbf{y} - \mathcal{A}\mathbf{x}_a\Big|\Big|_2^2 \\
&= \Big|\Big|\underbrace{\mathbf{y} - \mathcal{A}\mathbf{x}_a}_{\mathbf{v}} - \underbrace{\mathcal{A}\boldsymbol{\Delta}}_{\mathbf{z}}\Big|\Big|_2^2 - \Big|\Big|\underbrace{\mathbf{y} - \mathcal{A}\mathbf{x}_a}_{\mathbf{v}}\Big|\Big|_2^2 \\
&= \big(\mathbf{v} - \mathbf{z}\big)^\dagger\big(\mathbf{v} - \mathbf{z}\big) - \mathbf{v}^\dagger\mathbf{v} \\
&= \mathbf{v}^\dagger\mathbf{v} - \mathbf{v}^\dagger\mathbf{z} - \mathbf{z}^\dagger\mathbf{v} + \mathbf{z}^\dagger\mathbf{z} - \mathbf{v}^\dagger\mathbf{v} \\
&= \mathbf{z}^\dagger\mathbf{z} - \Big(\underbrace{\mathbf{v}^\dagger\mathbf{z}}_{\underbrace{(\mathbf{z}^\dagger\mathbf{v})^\dagger}_{a+bi}} + \underbrace{\mathbf{z}^\dagger\mathbf{v}}_{a-bi}\Big) \\
&= \mathbf{z}^\dagger\mathbf{z} - 2Re\big(\mathbf{z}^\dagger\mathbf{v}\big) \\
&= \mathbf{z}^\dagger\mathbf{z} - 2Re\big(\mathbf{z}^\dagger\mathbf{y} - \mathbf{z}^\dagger\underbrace{\mathcal{A}\mathbf{x}_a}_{\mathbf{w}}\big) \\
&= \mathbf{z}^\dagger\mathbf{z} - 2Re\big(\mathbf{z}^\dagger\mathbf{y} - \mathbf{z}^\dagger\mathbf{w}\big)
\end{aligned}$$

Computational cost for the evaluation of $\mathbf{z}$ could be saved if all possible values of $\boldsymbol{\Delta}$ were pre-calculated before entering the loop. Unfortunately this is not possible since $\boldsymbol{\Delta}$ is sampled from the proposal distribution. However it is still possible to pre-calculate $\mathbf{z} = \mathcal{A}\boldsymbol{\Delta}$ for a change of 1 in a single pixel only. If a certain pixel is changed at step k in the MCMC run, the according result for $\mathbf{z}$ for this single pixel will be

$$\mathbf{z}_k = \mathcal{A}\boldsymbol{\Delta}_k = \mathcal{A}v_k\mathbf{1}_\mathbf{s} = v_k\mathcal{A}\mathbf{1}_\mathbf{s}$$

where v_k is sampled from the proposal distribution and $\mathbf{1}_s$ is a sparse vector with one non-zero element with value 1 for the selected pixel. This is possible due to the linearity of the FFT in $\mathcal{A}$. Then any displacement v_k can be obtained by a simple scalar multiplication within the loop if $\mathcal{A}\mathbf{1}_\mathbf{s}$ has been pre-calculated and stored in a matrix Z outside the loop for each pixel.

A similar trick is used for computing $\mathbf{w}_k$ at MCMC-step k. $\mathbf{w}_0 := \mathcal{A}\mathbf{x}_0$ can be calculated once before entering the MCMC-loop and then just passed on from k to k+1 by a matrix addition with complexity N compared to a full evaluation of $\mathcal{A}\mathbf{x}_a$ which is of order $Nlog(N)$ when $\mathcal{A}$ represents some FFT.

$$\mathbf{w}_{k+1} := \mathcal{A}\mathbf{x}_{k+1} = \mathcal{A}(\mathbf{x}_k + \boldsymbol{\Delta}) = \mathbf{w}_k + \mathcal{A}\boldsymbol{\Delta} = \mathbf{w}_k + v\mathcal{A}\mathbf{1}_\mathbf{s} = \mathbf{w}_k + \mathbf{z}_k$$

It is important to note here that the addition must only be done if the new suggested value for the pixel has been accepted. These two tricks avoid an explicit usage of the operator $\mathcal{A}$ inside the MCMC-loop as long its linearity is provided.

6.2.2 Signal power prior (Tikhonov)

The acceptance ratio for a problem with signal power or Tikhonov prior is

$$Acc = e^{-\frac{1}{2}\Big(\Phi_0 + 2\alpha\Phi_T\Big)}$$

The first part is already known and for Φ_T the argumentation analogous to above.

$$\begin{aligned}
\Phi_T &= \mathbf{x}_n^\dagger B \mathbf{x}_n - \mathbf{x}_a^\dagger B \mathbf{x}_a \\
&= \big(\mathbf{x}_a - \boldsymbol{\Delta}\big)^\dagger B \big(\mathbf{x}_a - \boldsymbol{\Delta}\big) - \mathbf{x}_a^\dagger B \mathbf{x}_a \\
&= \mathbf{x}_a^\dagger B \mathbf{x}_a + \boldsymbol{\Delta}^\dagger B \mathbf{x}_a + \boldsymbol{\Delta}^\dagger B \boldsymbol{\Delta} + \mathbf{x}_a^\dagger B \boldsymbol{\Delta} - \mathbf{x}_a^\dagger B \mathbf{x}_a \\
&= \boldsymbol{\Delta}^\dagger B \boldsymbol{\Delta} + \big(\mathbf{x}_a^\dagger B \boldsymbol{\Delta}\big)^\dagger + \mathbf{x}_a^\dagger B \boldsymbol{\Delta} \\
&= \boldsymbol{\Delta}^\dagger B \boldsymbol{\Delta} + 2Re\big(\mathbf{x}_a^\dagger B \boldsymbol{\Delta}\big)
\end{aligned}$$

In the case a new proposal state $\mathbf{x}_n$ is generated for each pixel i independently, each $\boldsymbol{\Delta}$ is sparse vector with the i-th element being the only non-zero element. B is a given regularization matrix.

References

[1] Kenneth Shirley Andrew Gelman. Inference from simulations and monitoring convergence. In *Handbook of Markov Chain Monte Carlo*, Chapman & Hall/CRC Handbooks of Modern Statistical Methods. Chapman and Hall/CRC, 2011.

[2] Berkin Bilgic. Reconstruction algorithms for MRI, 2013.

[3] Michael R. Thompson E. Mark Haacke, Robert W. Brown and Ramesh Venkatesan. *Magnetic Resonance Imaging: Physical Principles and Sequence Design.* Wiley and Sons, 1995.

[4] Paul F. Fougere et al. *Maximum Entropy and Bayesian Methods*, volume 39 of *Fundamental Theories of Physics.* Kluwer Academic Publishers, Dordrecht, Boston, London, 1990.

[5] Von der Linden W. Evertz H.G. and Arrigoni E. Fortgeschrittene Quantenmechanik (Lecture notes), 2014.

[6] Von der Linden W. Evertz H.G. and Schachinger E. Computer Simulations (Lecture notes), 2009.

[7] Bretthorst Larry G. *Lecture Notes in Statistics.* Springer-Verlag, 1997.

[8] Jeffrey S. Rosenthal Gareth O. Roberts. Optimal scaling for various metropolis-hastings algorithms. 16:351–367, 2001.

[9] Richardson S and Spiegelhalter D J Gilks W R. *Markov Chain Monte Carlo in Practice.* Chapmann & Hall, 1996.

[10] Hakon Gudbjartsson and Samuel Patz. The rician distribution of noisy MRI data. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 34(6):910–914, December 1995.

[11] Jacques Hadamard. Sur les problemes aux derivees partielles et leur signification physique. pages 49–52, 1902.

[12] Per Christian Hansen. *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion.* Siam Society for Industrial and Applied Mathematics, Philadelphia, 1998.

[13] Jerome Idier. *Bayesian Approach to Inverse Problems.* Wiley, 2008.

[14] Florian Knoll. Constrained MR image reconstruction of undersampled data from multiple coils, 2011.

[15] Cecile Louchet. Variational and bayesian models for image denoising : From total variation towards non-local means, 2008.

[16] David G. Luenberger. *Optimization by Vector Space Methods*. Wiley Professional Paperback Series, New York, 1969.

[17] Kenneth S. Miller. Complex linear least squares, 1973.

[18] Stephen G. Odaibo. A quantum mechanical review of magnetic resonance imaging. *arXiv:1210.0946 [physics]*, October 2012. arXiv: 1210.0946.

[19] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger. SENSE: Sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine: Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 42(5):952–962, November 1999.

[20] Philip M. Robson, Aaron K. Grant, Ananth J. Madhuranthakam, Riccardo Lattanzi, Daniel K. Sodickson, and Charles A. McKenzie. Comprehensive quantification of signal-to-noise ratio and g-factor for image-based and k-space-based parallel imaging reconstructions. *Magnetic Resonance in Medicine: Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 60(4):895–907, October 2008.

[21] Samuli Siltanen. Inverse problems (Lecture notes MAT-52506), 2009.

[22] Stanislav Sykora. Antenna reciprocity theorem in magnetic resonance - A summary. `http://www.ebyte.it/library/educards/nmr/Nmr_AntennaTheorem.html`, 2007. [Online; accessed 30-July-2014].

[23] Dose V. Von der Linden W. and Von Toussaint U. *Bayesian Probability Theory: Applications in the Physical Sciences*. Cambridge University Press, 2014.