DELFT UNIVERSITY OF TECHNOLOGY

REPORT 18-04

CGME for general-form regularization with an application to low-field MRI

ISSN 1389-6520

Reports of the Delft Institute of Applied Mathematics

Delft 2018

CGME for general-form regularization with an application to low-field MRI

M.L. de Leeuw den Bouter, M.B. van Gijzen and R.F. Remis

December 18, 2018

Abstract

We generalize the CGME (Conjugate Gradient Minimal Error) algorithm to the weighted and regularized least squares problem. Analysis of the convergence of generalized CGME and CGLS shows that CGME can be expected to perform better for ill-conditioned regularization matrices.

Two different types of regularization are considered: an ℓ_1 penalty and an ℓ_2 penalty. The ℓ_1 problem is solved using Iterative Reweighted Least Squares, which leads to an ill-conditioned regularization matrix. The two methods are applied in a low-field MRI framework. The MRI physics in a low-field scanner are simulated to generate a noisy signal.

When an ℓ_1 penalty is used and iterative reweighted least squares is employed, GCGLS needs significantly more iterations to converge than GCGME. GCGME has a regularizing effect that leads to fewer artifacts in our simulations. This effect seems to be stronger when a lower number of CG iterations is used. These two observations indicate that GCGME is a very promising alternative to GCGLS.

1 Introduction

In the natural sciences, inverse problems are often ill-posed, and their solutions are very sensitive to perturbations in the data. An example of such an ill-posed problem is the reconstruction of an image based on a magnetic resonance imaging (MRI) signal in a low-field setting. In this case, the signal-to-noise ratio (SNR) is low and the noise has a severe impact on the solution. Regularization is required to limit the influence of the noise as much as possible. We are interested in developing algorithms that can be used in this low-field MRI setting.

The model is described by the following linear system:

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{n}, \tag{1}$$

where, in our case, $\mathbf{A}$ is a known matrix, $\mathbf{x}$ is the unknown image and $\mathbf{b}$ is the measured MRI signal, which is contamined by noise $\mathbf{n}$. We can attempt to solve for $\mathbf{x}$ by finding a solution to the least squares problem

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2. \tag{2}$$

This can be done by applying the conjugate gradient method introduced by Hestenes and Stiefel in 1952 [12] to the normal equations

$$\mathbf{A}^H\mathbf{A}\mathbf{x} = \mathbf{A}^H\mathbf{b}. \tag{3}$$

Here $\mathbf{A}^H$ denotes the Hermitian transpose of $\mathbf{A}$.

The conjugate gradient method tailored to equation (3) was proposed in [12] and is usually denoted by CGLS (Conjugate Gradient for Least Squares). The difference with the standard conjugate gradient method lies in the increased stability of the CGLS method. A review of the literature reveals that this method is known by other names as well. In [16], Saad calls it CGNR (Conjugate Gradient Normal Residual), while Hanke [8] and Engl [6] use the term CGNE (Conjugate Gradient for the Normal Equations).

On the other hand, the second normal equations

$$\mathbf{A}\mathbf{A}^H\mathbf{y} = \mathbf{b},\ \mathbf{x} = \mathbf{A}^H\mathbf{y} \tag{4}$$

can be solved using the conjugate gradient method as well. In literature, this is usually called CGME (Conjugate Gradient Minimal Error). However, in [2] it is called CGNE (Conjugate Gradient Normal Error), while [17] uses the term Craig's method. It was introduced by Craig in 1955 [5]. CGLS and CGME are discussed by Björck in [3], Hanke in [8] and Saad in [16]. While CGLS minimizes the residual $\mathbf{r} = \mathbf{b} - \mathbf{A}\mathbf{x}$ in the ℓ_2 norm over the Krylov subspace $\mathbf{x}_0 + \mathcal{K}_k(\mathbf{A}^H\mathbf{A}, \mathbf{A}^H\mathbf{b} - \mathbf{A}^H\mathbf{A}\mathbf{x}_0)$, CGME minimizes the error (over the same subspace). The main drawback of this method is that it only works for consistent problems $\mathbf{b} \in \mathbf{R}(\mathbf{A})$. This means that the method is of limited use for most problems in practice, because the presence of noise renders the system inconsistent. In [13], this problem is circumvented by defining an operator $\mathbf{Q}$ that projects $\mathbf{b}$ onto the column space of $\mathbf{A}$. Subsequently, $\mathbf{A}\mathbf{x} = \mathbf{Q}\mathbf{b}$ can be solved using CGME. The obvious disadvantage of this method is that $\mathbf{Q}\mathbf{b}$ has to be calculated and stored.

1.1 Regularization of the problem

Regularization of an ill-posed problem aims to make the problem less sensitive to noise by taking into account additional information. Many iterative methods have a regularizing effect if the iterating procedure is stopped early. The regularizing properties of CGLS are well known. They were established by Nemirovskii in [15] and are discussed in [3], [6] and [8], among others. CGME's regularizing effect was shown by Hanke in [9]. However, we are interested in what Hansen [11] calls general-form Tikhonov regularization, i.e. by adding a regularization term to minimization problem (2).

It is straightforward to generalize CGLS to regularized and weighted least squares problems of the form

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2_{\mathbf{W}} + \frac{1}{2}\tau\|\mathbf{x}\|^2_{\mathbf{R}} \tag{5}$$

where $\mathbf{W}$, a weighting matrix, and $\mathbf{R}$ are Hermitian positive definite matrices. In this paper we will use $\mathbf{W} = \mathbf{C}^{-1}$, where $\mathbf{C}$ is the covariance matrix of the noise. The optimal value of the regularization parameter τ is usually unknown. An approach that is often used to find a suitable value is the L-curve method [10]. By taking the gradient and setting it equal to 0, the normal equations are obtained:

$$(\mathbf{A}^H\mathbf{C}^{-1}\mathbf{A} + \tau\mathbf{R})\mathbf{x} = \mathbf{A}^H\mathbf{C}^{-1}\mathbf{b}. \tag{6}$$

Again, the conjugate gradient method can be used to solve equation (6). We will use the term GCGLS (Generalized CGLS) to refer to the conjugate gradient method applied to the normal equations (6).

Saunders [17] extended CGME to the regularized least squares problem with $\mathbf{R} = \mathbf{I}$ before. He introduces an additional variable $\mathbf{s}$ and considers the constrained minimization problem

$$\begin{aligned} &\min_{\mathbf{x},\mathbf{s}} \frac{1}{2} \left\| \begin{pmatrix} \mathbf{x} \\ \mathbf{s} \end{pmatrix} \right\|^2 \\ &\text{subject to } \begin{pmatrix} \mathbf{A} & \sqrt{\tau}\mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{s} \end{pmatrix} = \mathbf{b}. \end{aligned} \tag{7}$$

By defining $\mathbf{r} = \sqrt{\tau}\mathbf{s} = \mathbf{b} - \mathbf{A}\mathbf{x}$, he shows that this constrained minimization problem is equivalent to

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \frac{1}{2}\tau\|\mathbf{x}\|^2. \tag{8}$$

For every $\tau > 0$, $\begin{pmatrix} \mathbf{A} & \sqrt{\tau}\mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{s} \end{pmatrix} = \mathbf{b}$ is consistent and hence, equation (8) can be solved using CGME. Unfortunately, no advantages to using CGME were found. Note that such a reformulization is necessary because the standard way of including the regularization matrix $\mathbf{R} = \mathbf{I}$, by simply solving the so-called damped least squares problem

$$\begin{pmatrix} \mathbf{A} \\ \sqrt{\tau}\mathbf{I} \end{pmatrix} \mathbf{x} = \begin{pmatrix} \mathbf{b} \\ \mathbf{0} \end{pmatrix} \tag{9}$$

using CGME, is not possible, due to the inconsistency of the system.

In this paper, we reformulate the weighted and regularized least squares problem such that CGME can be used to solve it for nontrivial covariance and regularization matrices, filling a gap in existing literature as far as we know. We do this by deriving the Schur complement equation for the residual. A similar approach is taken by Arioli and Orban [1] to derive generalizations of the Golub-Kahan algorithm. Using these algorithms, they formulate generalizations of LSQR, Craig's method and LSMR for the general regularization problem. We explain in what cases Generalized CGME (GCGME) may have an advantage over GCGLS. Additionally, we apply GCGME to MRI data with different types of regularization.

1.2 Application to low-field MRI

This research is part of a project that aims towards creating an inexpensive MRI scanner using a Halbach cylinder that can be used for medical purposes. A Halbach cylinder is a configuration of permanent magnets that, combined, generate a magnetic field inside the cylinder and a very weak, or in the ideal case, no magnetic field outside of it. Imaging can be done by making use of the variations in the magnetic field. However, the resulting reconstruction problem is very ill-posed: if the variations are too small, the signal contains little spatial information, whereas if the variations are too high, the signal-to-noise ratio is the limiting factor. Nevertheless, in a similar project, Cooley et al. [4] have

shown that it is possible to reconstruct magnetic resonance images given signals obtained using a device based on a Halbach cylinder. The present paper results from our effort to address the challenges of low-field MRI using advanced image processing.

1.3 Structure

In section 2, a second set of normal equations is derived for the regularized and weighted least squares problem. The conjugate gradient method is tailored specifically to these equations, leading to the generalized CGME algorithm. In section 3, two different types of regularization are discussed and in section 4, we describe the application and the model that will be used to compare the performance of the two algorithms. In section 5, we describe how we obtained our dataset and results are given in section 6. Section 7 contains the conclusion.

2 GCGLS and GCGME

Minimization problem (5) can be formulated as a constrained minimization problem:

$$\begin{aligned} &\min_{\mathbf{r},\mathbf{x}} \frac{1}{2}||\mathbf{r}||_{\mathbf{C}}^2 + \frac{1}{2}\tau||\mathbf{x}||_{\mathbf{R}}^2 \\ &\text{s.t. } \mathbf{r} = \mathbf{C}^{-1}(\mathbf{b} - \mathbf{A}\mathbf{x}). \end{aligned} \tag{10}$$

Using the technique of Lagrange multipliers, we find that

$$\mathbf{r} = \mathbf{C}^{-1}(\mathbf{b} - \mathbf{A}\mathbf{x}) \text{ and } \tau\mathbf{R}\mathbf{x} = \mathbf{A}^H\mathbf{r}. \tag{11}$$

2.1 GCGLS

If we eliminate $\mathbf{r}$ from equation (11), the original normal equations (6) are obtained. By applying the conjugate gradient method to equation (6) and making some adjustments to increase stability (see [3] for details), the GCGLS algorithm is obtained:

Algorithm 1 GCGLS

Require: $\mathbf{A} \in \mathbf{C}^{M\times N}, \mathbf{C} \in \mathbf{C}^{M\times M}, \mathbf{R} \in \mathbf{C}^{N\times N}, \mathbf{x}_0, \in \mathbf{C}^N, \mathbf{b} \in \mathbf{C}^M, \tau \in \mathbf{R}_{\geq 0}$;
Ensure: Approximate solution $\mathbf{x}_k$ such that $\|\mathbf{A}^H\mathbf{r}_k - \tau\mathbf{R}\mathbf{x}_k\| \leqslant TOL$.

$\mathbf{r}_0 = \mathbf{C}^{-1}(\mathbf{b} - \mathbf{A}\mathbf{x}_0)$, $\mathbf{s}_0 = \mathbf{A}^H\mathbf{r}_0 - \tau\mathbf{R}\mathbf{x}_0$, $\mathbf{p}_0 = \mathbf{s}_0$; $\mathbf{q}_0 = \mathbf{A}^H\mathbf{p}_0$, $\gamma_0 = \mathbf{s}_0^H\mathbf{s}_0$, $k = 0$
while $\sqrt{\gamma_k} > TOL$ **and** $k < k_{max}$ **do**
  $\xi_k = \mathbf{q}_k^H\mathbf{C}^{-1}\mathbf{q}_k + \tau\mathbf{p}_k^H\mathbf{R}\mathbf{p}_k$
  $\alpha_k = \frac{\gamma_k}{\xi_k}$
  $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k\mathbf{p}_k$
  $\mathbf{R}\mathbf{x}_{k+1} = \mathbf{R}\mathbf{x}_k + \alpha_k\mathbf{R}\mathbf{p}_k$
  $\mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_k\mathbf{C}^{-1}\mathbf{q}_k$
  $\mathbf{s}_{k+1} = \mathbf{A}^H\mathbf{r}_{k+1} - \tau\mathbf{R}\mathbf{x}_{k+1}$
  $\gamma_{k+1} = \mathbf{s}_{k+1}^H\mathbf{s}_{k+1}$
  $\beta_k = \frac{\gamma_{k+1}}{\gamma_k}$
  $\mathbf{p}_{k+1} = \mathbf{s}_{k+1} + \beta_k\mathbf{p}_k$
  $\mathbf{q}_{k+1} = \mathbf{A}\mathbf{p}_{k+1}$
  $k = k + 1$
end while

Here, $\mathbf{s}_k$ denotes the residual of the normal equations (6). We remark that the vectors on the left side can be overwritten by the vectors on the right. Only 8 vectors have to be stored: $\mathbf{x}$, $\mathbf{r}$, $\mathbf{s}$, $\mathbf{p}$, $\mathbf{q}$, $\mathbf{R}\mathbf{x}$, $\mathbf{R}\mathbf{p}$ and $\mathbf{C}^{-1}\mathbf{q}$. Note that the recursion for $\mathbf{R}\mathbf{x}_{k+1}$ is included to avoid an extra multiplication with $\mathbf{R}$. It can be ignored in case $\mathbf{R} = \mathbf{I}$. In this algorithm, only three matrix-vector multiplications are carried out per iteration: $\mathbf{A}\mathbf{p}_{k+1}$, $\mathbf{A}^H\mathbf{r}_k$ and $\mathbf{R}\mathbf{p}_k$. Additionally, one system with $\mathbf{C}$ has to be solved (if $\mathbf{C} \neq \mathbf{I}$). A slightly different formulation of the GCGLS algorithm can be found in [18].

2.2 GCGME

If $\tau\mathbf{R}$ is invertible, $\mathbf{x}$ can be eliminated from equation (11), yielding

$$\left(\frac{1}{\tau}\mathbf{A}\mathbf{R}^{-1}\mathbf{A}^H + \mathbf{C}\right)\mathbf{r} = \mathbf{b}. \tag{12}$$

Subsequently, $\mathbf{x}$ can be obtained from $\mathbf{r}$:

$$\mathbf{x} = \frac{1}{\tau}\mathbf{R}^{-1}\mathbf{A}^H\mathbf{r}. \tag{13}$$

In [1], Arioli and Orban derive a generalization of Craig's method [5] based on Schur complement (12). Below we formulate a similar generalization of the CGME method applied to this system. We are not aware this generalization of CGME has been formulated elsewhere.

Algorithm 2 GCGME

Require: $\mathbf{A} \in \mathbf{C}^{M\times N}, \mathbf{C} \in \mathbf{C}^{M\times M}, \mathbf{R} \in \mathbf{C}^{N\times N}, \mathbf{r}_0 \in \mathbf{C}^M, \mathbf{b} \in \mathbf{C}^M, \tau \in \mathbf{R}_{>0}$;
Ensure: Approximate solution $\mathbf{x}_k$ such that $\|\mathbf{b} - \mathbf{A}\mathbf{x}_k - \mathbf{C}\mathbf{r}_k\| \leqslant TOL$.

$\mathbf{x}_0 = \frac{1}{\tau}\mathbf{R}^{-1}\mathbf{A}^H\mathbf{r}_0$
$\mathbf{s}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0 - \mathbf{C}\mathbf{r}_0$, $\mathbf{p}_0 = \mathbf{s}_0$, $\mathbf{q}_0 = \mathbf{A}^H\mathbf{p}_0$, $\gamma_0 = \mathbf{s}_0^H\mathbf{s}_0$, $k = 0$
while $\sqrt{\gamma_k} > TOL$ **and** $k < k_{max}$ **do**
  $\xi_k = \frac{1}{\tau}\mathbf{q}_k^H\mathbf{R}^{-1}\mathbf{q}_k + \mathbf{p}_k^H\mathbf{C}\mathbf{p}_k$
  $\alpha_k = \frac{\gamma_k}{\xi_k}$
  $\mathbf{r}_{k+1} = \mathbf{r}_k + \alpha_k\mathbf{p}_k$
  $\mathbf{x}_{k+1} = \mathbf{x}_k + \frac{\alpha_k}{\tau}\mathbf{R}^{-1}\mathbf{q}_k$
  $\mathbf{s}_{k+1} = \mathbf{s}_k - \alpha_k(\frac{1}{\tau}\mathbf{A}\mathbf{R}^{-1}\mathbf{q}_k + \mathbf{C}\mathbf{p}_k)$
  $\gamma_{k+1} = \mathbf{s}_{k+1}^H\mathbf{s}_{k+1}$
  $\beta_k = \frac{\gamma_{k+1}}{\gamma_k}$
  $\mathbf{p}_{k+1} = \mathbf{s}_{k+1} + \beta_k\mathbf{p}_k$
  $\mathbf{q}_{k+1} = \mathbf{A}^H\mathbf{p}_{k+1}$
  $k = k + 1$
end while

Here, $\mathbf{s}_k$ is the residual of the normal equations (12). Note that the original CGME algorithm can be recovered from the generalized CGME algorithm given above by taking $\frac{1}{\tau}\mathbf{R} = \mathbf{I}$ and $\mathbf{C} = \mathbf{O}$. Only 7 vectors have to be stored: $\mathbf{x}$, $\mathbf{r}$, $\mathbf{s}$, $\mathbf{p}$, $\mathbf{q}$, $\mathbf{R}^{-1}\mathbf{q}$ and $\mathbf{C}\mathbf{p}$. Like GCGLS, GCGME needs four matrix operations per iteration: $\mathbf{C}\mathbf{p}_k$, $\mathbf{A}^H\mathbf{p}_k$, $\mathbf{R}^{-1}\mathbf{q}_k$ and $\mathbf{A}\mathbf{R}^{-1}\mathbf{q}_k$.

We remark that there is an essential difference between GCGLS and GCGME. GCGLS iterates for the solution vector $\mathbf{x}$ and the equality $\mathbf{r}_k = \mathbf{C}^{-1}(\mathbf{b}-\mathbf{A}\mathbf{x}_k)$ is explicitly imposed. The equality $\mathbf{x}_k = \frac{1}{\tau}\mathbf{R}^{-1}\mathbf{A}^H\mathbf{r}_k$ is not enforced, and is only (approximately) satisfied after convergence. GCGME, on the other hand, iterates for $\mathbf{r}_k$. The equality $\mathbf{x}_k = \frac{1}{\tau}\mathbf{R}^{-1}\mathbf{A}^H\mathbf{r}_k$ is enforced, while $\mathbf{r}_k = \mathbf{C}^{-1}(\mathbf{b}-\mathbf{A}\mathbf{x}_k)$ is only satisfied approximately after convergence.

2.3 Analysis of GCGLS and GCGME

2.3.1 Optimality property

The conjugate gradient method minimizes the error in the norm induced by the system matrix. This means that, in every iteration, GCGLS minimizes

$$\begin{aligned}
&\min_{\mathbf{x}_k}(\mathbf{x}_k - \mathbf{x})^H(\mathbf{A}^H\mathbf{C}^{-1}\mathbf{A} + \tau\mathbf{R})(\mathbf{x}_k - \mathbf{x}) \Leftrightarrow \qquad (14)\\
&\min_{\mathbf{x}_k}(\mathbf{r} - \mathbf{r}_k)^H\mathbf{C}(\mathbf{r} - \mathbf{r}_k) + \tau(\mathbf{x}_k - \mathbf{x})^H\mathbf{R}(\mathbf{x}_k - \mathbf{x}) \ , \quad \mathbf{r}_k = \mathbf{C}^{-1}(\mathbf{b} - \mathbf{A}\mathbf{x}_k) \Leftrightarrow\\
&\min_{\mathbf{x}_k}\|\mathbf{r} - \mathbf{r}_k\|_{\mathbf{C}}^2 + \tau\|\mathbf{x} - \mathbf{x}_k\|_{\mathbf{R}}^2 \ , \quad \mathbf{r}_k = \mathbf{C}^{-1}(\mathbf{b} - \mathbf{A}\mathbf{x}_k)
\end{aligned}$$

for $\mathbf{x}_k - \mathbf{x}_0 \in \mathcal{K}_k(\mathbf{A}^H\mathbf{C}^{-1}\mathbf{A} + \tau\mathbf{R}, \mathbf{A}^H\mathbf{C}^{-1}\mathbf{r}_0 + \tau\mathbf{R}\mathbf{x}_0)$. For every iteration of GCGME, the following holds:

$$\begin{aligned}
&\min_{\mathbf{r}_k}(\mathbf{r} - \mathbf{r}_k)^H(\frac{1}{\tau}\mathbf{A}\mathbf{R}^{-1}\mathbf{A}^H + \mathbf{C})(\mathbf{r} - \mathbf{r}_k) \Leftrightarrow \qquad (15)\\
\min_{\mathbf{r}_k}\tau(\mathbf{x} - \mathbf{x}_k)^H\mathbf{R}(\mathbf{x} - \mathbf{x}_k) + (\mathbf{r} - \mathbf{r}_k)^H\mathbf{C}(\mathbf{r} - \mathbf{r}_k), \quad &\mathbf{x}_k = \frac{1}{\tau}\mathbf{R}^{-1}\mathbf{A}^H\mathbf{r}_k\\
\min_{\mathbf{r}_k}\|\mathbf{r} - \mathbf{r}_k\|_{\mathbf{C}}^2 + \tau\|\mathbf{x} - \mathbf{x}_k\|_{\mathbf{R}}^2 \ , \quad &\mathbf{x}_k = \frac{1}{\tau}\mathbf{R}^{-1}\mathbf{A}^H\mathbf{r}_k
\end{aligned}$$

with $\mathbf{r}_k - \mathbf{r}_0 \in \mathcal{K}_k(\mathbf{r}_0, \frac{1}{\tau}\mathbf{A}\mathbf{R}^{-1}\mathbf{A}^H + \mathbf{C})$. Note that GCGLS and GCGME minimize the same weighted combination of the errors in the residual and in the solution, but over different subspaces and under different constraints.

2.3.2 Convergence

The convergence of the conjugate gradient method depends on the condition number of the system matrix. Suppose CG is used to solve the system $\mathbf{K}\mathbf{u} = \mathbf{f}$ for the unknown vector $\mathbf{u}$, where $\mathbf{K}$ is a Hermitian positive definite (HPD) matrix and $\mathbf{f}$ is a known vector. Then the following classical convergence bound holds [3]:

$$\|\mathbf{u} - \mathbf{u}_k\|_{\mathbf{K}} \leq 2\left(\frac{\sqrt{\kappa_2(\mathbf{K})} - 1}{\sqrt{\kappa_2(\mathbf{K})} + 1}\right)^k \|\mathbf{u} - \mathbf{u}_0\|_{\mathbf{K}}, \tag{16}$$

where $\kappa_2(\mathbf{K})$ is the ℓ_2 norm condition number of $\mathbf{K}$, which, for HPD matrices, is equal to

$$\kappa_2(\mathbf{K}) = \frac{\lambda_{\max}(\mathbf{K})}{\lambda_{\min}(\mathbf{K})} \tag{17}$$

in which $\lambda_{\max}(\mathbf{K})$ and $\lambda_{\min}(\mathbf{K})$ are the largest and smallest eigenvalue of $\mathbf{K}$, respectively. In this section we bound the condition numbers of the two Schur complement matrices in equations (6) and (12) to gain insight into when GCGME can be expected to perform better than GCGLS, and vice versa. Given two HPD matrices $\mathbf{K}$ and $\mathbf{M}$, the following bound on the condition number holds:

$$\frac{\lambda_{\max}(\mathbf{K}) + \lambda_{\min}(\mathbf{M})}{\lambda_{\min}(\mathbf{K}) + \lambda_{\max}(\mathbf{M})} \leq \kappa_2(\mathbf{K} + \mathbf{M}) \leq \frac{\lambda_{\max}(\mathbf{K}) + \lambda_{\max}(\mathbf{M})}{\lambda_{\min}(\mathbf{K}) + \lambda_{\min}(\mathbf{M})}. \tag{18}$$

This inequality can be proved using Weyl's theorem [21], which states that for eigenvalues of Hermitian matrices $\mathbf{K}$ and $\mathbf{M}$ the following holds:

$$\lambda_i(\mathbf{K}) + \lambda_{\min}(\mathbf{M}) \leq \lambda_i(\mathbf{K} + \mathbf{M}) \leq \lambda_i(\mathbf{K}) + \lambda_{\max}(\mathbf{M}). \tag{19}$$

Here $\lambda_i(\mathbf{K})$ denotes any eigenvalue of the matrix $\mathbf{K}$. For GCGLS we have that

$$\mathbf{K} = \tau\mathbf{R} \quad , \quad \mathbf{M} = \mathbf{A}^H\mathbf{C}^{-1}\mathbf{A} \tag{20}$$

and, using the following inequalities

$$\lambda_{\max}(\mathbf{A}^H\mathbf{C}^{-1}\mathbf{A}) \leq \frac{\sigma_{\max}(\mathbf{A})^2}{\lambda_{\min}(\mathbf{C})} \ , \quad \lambda_{\min}(\mathbf{A}^H\mathbf{C}^{-1}\mathbf{A}) \geq 0, \tag{21}$$

with $\sigma_{\max}(\mathbf{A})$ the largest singular value of $\mathbf{A}$, we get that

$$\kappa_2(\mathbf{A}^H\mathbf{C}^{-1}\mathbf{A} + \tau\mathbf{R}) \geq \frac{\tau\lambda_{\max}(\mathbf{R})\lambda_{\min}(\mathbf{C})}{\tau\lambda_{\min}(\mathbf{R})\lambda_{\min}(\mathbf{C}) + \sigma_{\max}(\mathbf{A})^2} \tag{22}$$

Analogously we get for GCGME that

$$\kappa_2(\mathbf{A}\mathbf{R}^{-1}\mathbf{A}^H + \tau\mathbf{C}) \geq \frac{\tau\lambda_{\max}(\mathbf{C})\lambda_{\min}(\mathbf{R})}{\tau\lambda_{\min}(\mathbf{R})\lambda_{\min}(\mathbf{C}) + \sigma_{\max}(\mathbf{A})^2} \quad . \tag{23}$$

These inequalities indicate that if

$$\lambda_{\max}(\mathbf{C})\lambda_{\min}(\mathbf{R}) \gg \lambda_{\max}(\mathbf{R})\lambda_{\min}(\mathbf{C}) \Leftrightarrow \kappa_2(\mathbf{C}) \gg \kappa_2(\mathbf{R}) \tag{24}$$

GCGLS can be expected to perform best, and vice versa that if

$$\kappa_2(\mathbf{R}) \gg \kappa_2(\mathbf{C}) \tag{25}$$

GCGME should be preferred.

3 Types of regularization

In this research, we focus on ℓ_1 and ℓ_2 penalty terms for regularization. When an ℓ_2 penalty is used, the resulting minimization problem is

$$\min_{\mathbf{x}} \frac{1}{2}||\mathbf{Ax}-\mathbf{b}||_2^2 + \frac{1}{2}\tau||\mathbf{Fx}||_2^2. \tag{26}$$

This is simply equation (5) with $\mathbf{R} = \mathbf{F}^H\mathbf{F}$. We can solve it using either GCGLS or GCGME. More generally speaking, in the case of an ℓ_p penalty, the minimization problem becomes

$$\min_{\mathbf{x}} \frac{1}{2}||\mathbf{Ax}-\mathbf{b}||_2^2 + \frac{1}{2}\tau||\mathbf{Fx}||_p^p, \tag{27}$$

with $p \in (0, 2]$. In both cases, $\mathbf{F}$ is some regularizing matrix. Note that equation (27) can be rewritten as

$$\min_{\mathbf{x}} \frac{1}{2}||\mathbf{Ax}-\mathbf{b}||_2^2 + \frac{1}{2}\tau||\mathbf{x}||_{\mathbf{F}^H\mathbf{DF}}^2, \tag{28}$$

where

$$\mathbf{D} := \operatorname{diag}\left(\frac{1}{|\mathbf{Fx}|^{2-p}}\right), \tag{29}$$

where $|\mathbf{Fx}|$ is simply the vector $\mathbf{Fx}$ in absolute value. Minimization problem (28) can be solved using Iterative Reweighted Least Squares (IRLS), see for example [3]. So, when the kth iterate $\mathbf{x}_k$ is known, $\mathbf{x}_{k+1}$ is found as follows:

$$\mathbf{x}_{k+1} = \arg\min_{\mathbf{x}} \frac{1}{2}||\mathbf{Ax}-\mathbf{b}||_2^2 + \frac{1}{2}\tau||\mathbf{x}||_{\mathbf{F}^H\mathbf{D}_k\mathbf{F}}^2, \tag{30}$$

where

$$\mathbf{D}_k = \operatorname{diag}\left(\frac{1}{|\mathbf{Fx}_k|^{2-p}+\epsilon}\right). \tag{31}$$

This is repeated until convergence. Furthermore, ϵ is a small number that is added to the denumerator to prevent division by zero. We will use $\epsilon = 10^{-6}$. When carrying out calculations with $\mathbf{D}_k^{-1}$, we will use

$$\mathbf{D}_k^{-1} = \operatorname{diag}\left(|\mathbf{Fx}_k|^{2-p}\right). \tag{32}$$

Due to the sparsity-inducing property of the ℓ_1 penalty, $\mathbf{D}_k^{-1} = \operatorname{diag}(|\mathbf{Fx}_k|)$ will contain an increasing amount of entries that are equal to zero. In case $\mathbf{F}$ is

an invertible matrix, $\mathbf{R}_k^{-1} = \mathbf{F}^{-1}\mathbf{D}_k^{-1}(\mathbf{F}^H)^{-1}$. When GCGME is used, we can take advantage of this structure, instead of calculating $\mathbf{R}_k$ and working with its inverse. When $\mathbf{F}$ is an orthogonal matrix, no additional computations are necessary to compute inverses.

The regularization matrix $\mathbf{R} = \mathbf{F}^H\mathbf{D}_k\mathbf{F}$ will become ill-conditioned when elements of $\mathbf{F}\mathbf{x}_k$ become small (and $p < 2$). Therefore, we expect that, when combined with IRLS, GCGME will perform better than GCGLS. Numerical experiments are carried out to investigate this.

4 Application: MRI

4.1 Modeling the signal

In magnetic resonance imaging (MRI), the internal structure of the body is made visible by measuring a voltage signal that is induced by time variations of the transverse magnetization within a body part of interest. Based on this measured signal, an image of the spin density ρ of different tissue types may be obtained.

To be specific, first the body part of interest is placed in a static magnetic field $\vec{B} = B_0(\vec{r})\,\vec{i}_x$ that is oriented in the x-direction in our Halbach measurement setup (see figure 1) with a position dependent x-component $B_0 = B_0(\vec{r})$. A net magnetization

$$\vec{M}_{\text{eq}} = M_0(\vec{r})\,\vec{i}_x \quad \text{with} \quad M_0(\vec{r}) = \frac{\gamma^2\hbar^2}{4k_{\text{B}}T}\rho(\vec{r})B_0(\vec{r}) \tag{33}$$

will be induced that is oriented in the same direction as the static magnetic field. In the above expression, $\gamma = 267 \cdot 10^6$ rad s^{-1} T^{-1} is the proton gyromagnetic ratio, $\hbar = 1.055 \cdot 10^{-34}$ m^2 kg s^{-1} is Planck's constant divided by 2π, $k_{\text{B}} = 1.381 \cdot 10^{-23}$ m^2 kg s^{-2} K^{-1} is Boltzmann's constant, and T is the temperature in kelvin.

Subsequently, a radiofrequency pulse is emitted to tip the magnetization towards the transverse yz-plane. After this pulse has been switched off (in our model at $t = 0$), the magnetization rotates about the static magnetic field with a precessional frequency (or Larmor frequency)

$$\omega(\vec{r}) = \gamma B_0(\vec{r}) \tag{34}$$

and will relax back to its equilibrium given by equation (33). During this process, an electromagnetic field is generated that can be locally measured outside the body using a receive coil. This measured signal is amplified, demodulated, and low-pass filtered and for the resulting signal we have the signal model [14]:

$$S(t) = \int_{\vec{r}\in\mathbb{D}} c(\vec{r})\omega(\vec{r})e^{-t/T_2(\vec{r})}M_\perp(\vec{r},0)e^{-\mathrm{i}\gamma\Delta\omega t}\,\mathrm{d}\,\vec{r}, \tag{35}$$

where $\mathbb{D}$ is the domain occupied by the body part of interest, $T_2(\vec{r})$ is the transverse relaxation time, $c(\vec{r})$ is the so-called coil sensitivity with amplification included, $M_\perp(\vec{r},0)$ is the transverse magnetization at $t = 0$, and $\Delta\omega$ is the difference between the Larmor frequency and the demodulation frequency

that is used. We will set this demodulation frequency to be equal to the frequency that corresponds to the static magnetic field at the center of our imaging domain.

Furthermore, using equation (34) in the expression for M_0, we have

$$M_0(\vec{r}) = \frac{\gamma\hbar^2}{4k_{\mathrm{B}}T}\rho(\vec{r})\omega(\vec{r}) \tag{36}$$

and since the initial transverse magnetization $M_\perp(\vec{r}, 0)$ is proportional to $M_0(\vec{r})$, we can also write our signal model as

$$S(t) = \int_{\vec{r}\in\mathbb{D}} c(\vec{r})\omega^2(\vec{r})e^{-t/T_2(\vec{r})}\rho(\vec{r})e^{-\mathrm{i}\gamma\Delta\omega t}\,\mathrm{d}\vec{r}, \tag{37}$$

where it is understood that all remaining proportionality constants have been incorporated in the coil sensitivity $c(\vec{r})$. Finally, we note that in low-field MRI, signal decay is dominated by spatial variations in the static background field and transverse T_2 signal decay may be ignored. Taking this observation into account, our final signal model becomes

$$S(t) = \int_{\vec{r}\in\mathbb{D}} c(\vec{r})\omega^2(\vec{r})\rho(\vec{r})e^{-\mathrm{i}\gamma\Delta\omega t}\,\mathrm{d}\vec{r}. \tag{38}$$

The measurements carried out in an MRI scanner consist of noisy samples of the signal given by equation (38):

$$b_i = S(t_i) + e_i, \quad i = 1, ..., L, \tag{39}$$

where b_i denotes the ith sample of the signal, measured at time t_i. L is the number of time samples, e_i is the measurement error. In high-field MRI, the magnetic field is manipulated in such a way that equation (38) constitutes a Fourier Transform. The resulting linear problem is well posed, and the image can be efficiently obtained using an inverse FFT. However, in low-field MRI, the magnetic field is very inhomogeneous. This is an example of a growing number of instances where the FFT is inadequate. Model-based image reconstruction can be applied instead [7].

In order to estimate $\rho(\vec{r})$, a finite series expansion is used:

$$\rho(\vec{r}) = \sum_{j=1}^{N} x_j\phi(\vec{r} - \vec{r}_j), \tag{40}$$

where $\phi(\cdot)$ denotes the object basis function and $\vec{r}_j$ is the center of the jth basis function. x_j are the coefficients. Usually, rectangular basis functions are used. In that case, N is the number of pixels. Combining equations (38) and (40) yields

$$S(t_i) = \sum_{j=1}^{N} a_{ij}x_j, \tag{41}$$

where

$$a_{ij} = \int_{\mathrm{object}} \phi(\vec{r} - \vec{r}_j)c(\vec{r})\omega(\vec{r})^2 e^{-i\Delta\omega(\vec{r})t_i}\; d\vec{r}. \tag{42}$$

When the basis functions are highly localized, a "center of pixel" approximation can be used:

$$a_{ij} = c(\vec{r}_j)\omega(\vec{r}_j)^2 e^{-i\Delta\omega(\vec{r}_j)t_i}\Delta x\Delta y\Delta z. \tag{43}$$

Here, $\Delta x\Delta y$ is the pixel size and Δz is the thickness of the slice that is being imaged. Combining equations (39) and (40) yields one system of equations:

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{e}, \tag{44}$$

where the elements of $\mathbf{A}$ are described by equation (43). This problem is ill-posed due to the nature of the magnetic field that is used. As can be seen in figure 2, the field has a high degree of symmetry. The precessional frequency depends linearly on the magnitude of the field, which means that several pixels will correspond to the same frequency. Therefore, using only one measured signal, it is impossible to determine the contribution of each pixel to the signal. By rotating the object to be imaged and hence obtaining a multitude of different signals corresponding to different rotations of the same object, we plan to mitigate this problem. The same approach was taken by Cooley et al. [4].

4.2 Different choices for F

We will minimize the following two expressions to obtain approximations to the optimal solution $\mathbf{x}$:

$$\min_{\mathbf{x}} \frac{1}{2}||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2 + \frac{1}{2}\tau||\mathbf{F}\mathbf{x}||_2^2 \tag{45}$$

and

$$\min_{\mathbf{x}} \frac{1}{2}||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2 + \frac{1}{2}\tau||\mathbf{F}\mathbf{x}||_1. \tag{46}$$

Two different cases will be considered.

4.2.1 The identity as the regularization operator

First, we set $\mathbf{F} = \mathbf{I}$. When the penalty term is of the ℓ_2 form, this is standard Tikhonov regularization [20]. In case the ℓ_1 penalty is used, this is known as LASSO regularization which was first introduced by Tibshirani in [19]. If the regularization parameter is set to a sufficiently high value, the resulting solution will be sparse. The rationale behind choosing this type of regularization is the fact that the intensity of many pixels in MRI images is equal to 0. Since $\mathbf{F}$ is square, we do not have to explicitly calculate $\mathbf{R}$ to carry out calculations with $\mathbf{R}^{-1}$ when using GCGME.

4.2.2 Regularizing using first order differences

Additionally, we consider the case where $\mathbf{F}$ is a first-order difference matrix that calculates the values of the jumps between each pair of neighboring pixels:

$$\begin{aligned}||\mathbf{F}\mathbf{x}||_1 = &\sum_{k=1}^{n}\sum_{l=1}^{m-1}|X_{l,k} - X_{l+1,k}| + \sum_{k=1}^{n}|X_{m,k}| \\ &+ \sum_{l=1}^{m}\sum_{k=1}^{n-1}|X_{l,k} - X_{l,k+1}| + \sum_{l=1}^{m}|X_{l,n}|.\end{aligned} \tag{47}$$

The first term and the third sum over all jumps in the image. The other two terms sum over all pixels in the final row and the final column of the image. By doing this, we impose a Neumann boundary condition. We do this because we need $\mathbf{F}^H\mathbf{F}$ to be invertible. Here, X is the image, consisting of $m \times n$ pixels. $\mathbf{x}$ is X in vector form. This is known as total variation regularization. A reason for defining $\mathbf{F}$ like this is that neighboring pixels are very likely to have the same values in MR images. This is due to the fact that neighboring pixels tend to represent the same tissue. However, $\mathbf{F}$ is not a square matrix, which means that, in the ℓ_1 case, $\mathbf{R}_k$ has to be calculated explicitly and then inverted when GCGME is used. Although this makes regularization with first order differences in combination with GCGME less attractive than with GCGLS, we do include this technique to investigate the relative reconstruction quality of this widely used regularization method.

5 Experiments

The signal generation inside a Halbach cylinder is simulated using equations (43) and (44). Beforehand, the magnetic field inside a Halbach cylinder was simulated. Figure 1 shows the Halbach cylinder.

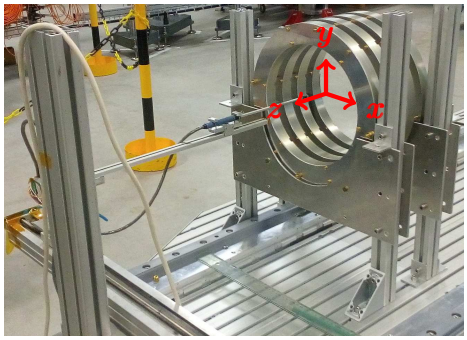


Figure 1: The Halbach cylinder.

The Field of View (FoV) is set to 14 cm in both directions. The field within the FoV at $z = 0$ is shown in figure 2. We assume a slice thickness of $\Delta z = 5mm$.

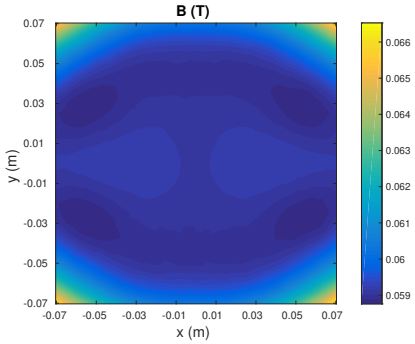
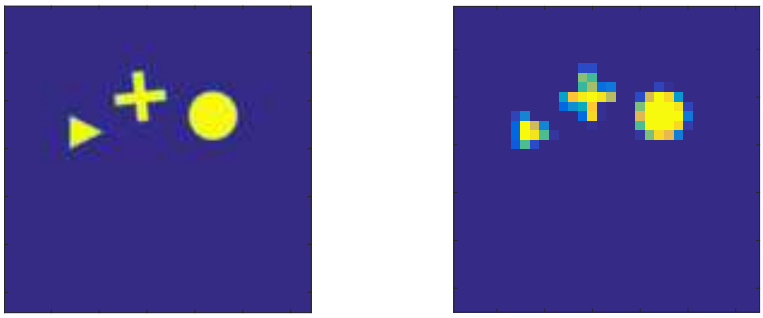


Figure 2: The magnetic field (in T) within the FoV.

The time step is set to $\Delta t = 5 \cdot 10^{-6}$, the number of time steps to $L = 101$ so for each pulse, the duration is 0.5 ms. Additionally, the field is rotated by 5° after each pulse. 72 different angles are considered. As a result, the system consists of 7272 equations.

The phantom is shown in figure 3. We are interested in finding a 32x32 pixel approximation of this phantom. However, we use a phantom of 128x128 pixels to generate the signal in order to obtain a more realistic signal.



(a) Phantom used to generate the signal. (b) Model solution.

Figure 3: The phantom.

The coil sensitivity c is assumed to be constant. Hence, it is left out of the calculations. White Gaussian noise with standard deviation 0.001 is added. Therefore, $\mathbf{C} = \mathbf{I}$.

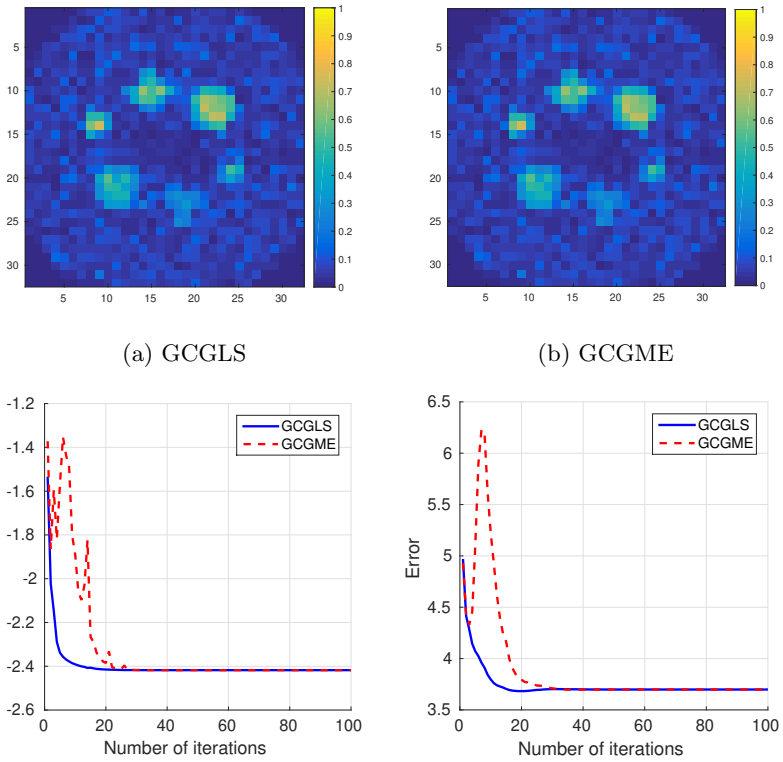
6 Results

The numerical experiments were carried out using MATLAB version 2015a.

6.1 ℓ_2 penalty

6.1.1 $\mathbf{R} = \mathbf{I}$

GCGLS and GCGME are applied to minimization problem (45). We first consider standard Tikhonov regularization, so $\mathbf{R} = \mathbf{I}$. The regularization parameter is chosen heuristically and is set to $\tau = 5 \cdot 10^{-5}$. The number of CG iterations is 100 for both GCGLS and GCGME. The resulting images are shown in figures 4a and 4b. Both types of CG lead to the same result, as expected and the time per iteration is the same (0.023 seconds). However, when using GCGME, more iterations are required before convergence is attained, as can be seen in figure 4c.



(a) GCGLS

(b) GCGME

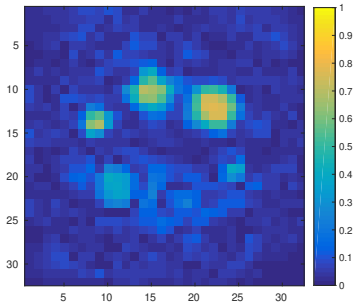
(c) Value of the objective function

(d) Error

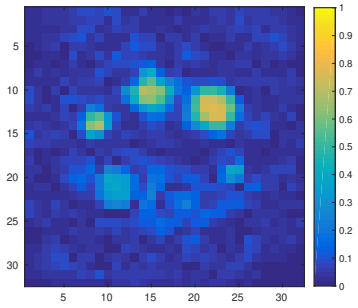
Figure 4: Results obtained using the GCGLS and GCGME algorithms. Here, standard Tikhonov regularization is used with $\tau = 5 \cdot 10^{-5}$. The value of objective function (5) and the error are measured in the ℓ_2 norm.

6.1.2 $\mathbf{R} = \mathbf{F}^H\mathbf{F}$

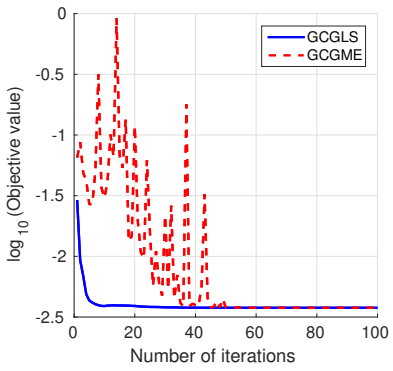
Now, $\mathbf{R}$ is defined as $\mathbf{R} = \mathbf{F}^H\mathbf{F}$, where $\mathbf{F}$ is the total variation operator. τ is set to $5 \cdot 10^{-5}$. Again, this value is chosen heuristically. The results are shown in figure 5. As expected, the two algorithms yield the same image. It is clear that GCGLS converges more rapidly than GCGME. One iteration of GCGLS is only slightly faster than a GCGME iteration (0.023 vs 0.024 seconds on average).



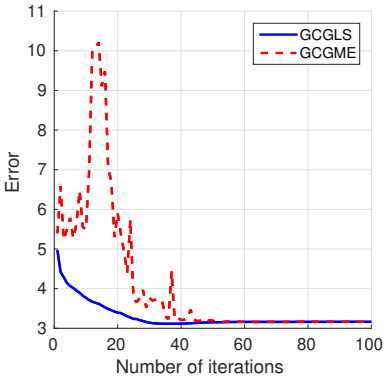
(a) GCGLS



(b) GCGME



(c) Value of the objective function



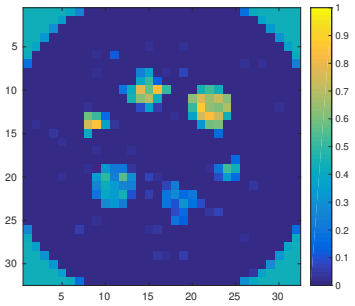
(d) Error

Figure 5: Results obtained using the GCGLS and GCGME algorithms. Here, Tikhonov regularization with $\mathbf{R} = \mathbf{F}^H\mathbf{F}$ is used with $\tau = 5 \cdot 10^{-5}$. The value of objective function (5) and the error are measured in the ℓ_2 norm.

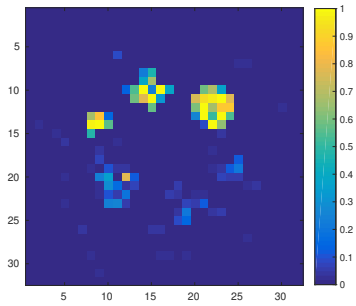
6.2 ℓ_1 penalty

6.2.1 LASSO

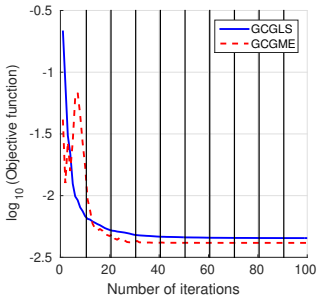
Next, we solve minimization problem (46) with $\mathbf{F} = \mathbf{I}$. The regularization parameter is set to $\tau = 3 \cdot 10^{-5}$. The number of IRLS iterations is set to 10. In each of the IRLS iterations, minimization problem (30) is solved using GCGLS or GCGME. In each IRLS iteration, the CG method is initialized with a different starting vector: the final iterate of the previous fixed-point iteration. The number of CG iterations is set to 10. As for the ℓ_2 case, both GCGLS and GCGME need 0.023 seconds per iteration. Figures 6a and 6b show the resulting images. Clearly, the GCGME result resembles the model solution much more than the GCGLS result does. This is reflected in figures 6c and 6d, which contain the plots showing the objective value and the error as a function of the iteration number.



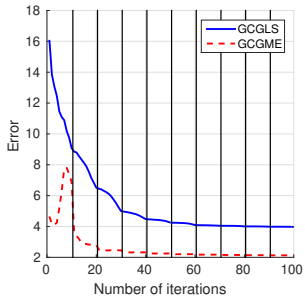
(a) GCGLS



(b) GCGME



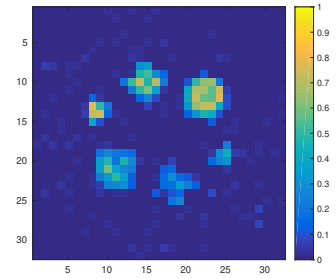
(c) Value of the objective function



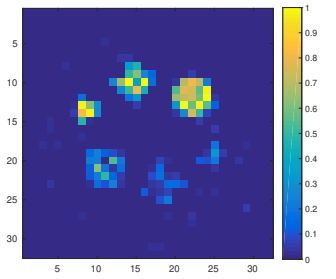
(d) Error

Figure 6: Results obtained using the GCGLS and GCGME algorithms when LASSO regularization is used. The regularization parameter is set to $\tau = 3 \cdot 10^{-5}$. The value of the objective function (46) and the error are measured in the ℓ_2 norm. The number of CG iterations is set to 10. In the plots, the vertical black lines indicate the start of a new fixed-point iteration.

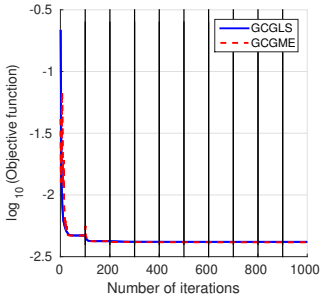
Increasing the number of CG iterations to 100 changes the results, see figure 7. We see that the GCGME image has not changed much (interestingly, it seems to look slightly worse than before), whereas the GCGLS one has. It is starting to resemble the GCGME result more, but it is still of inferior quality. Increasing the number of CG iterations to 1000 does not change the results significantly, see the appendix.



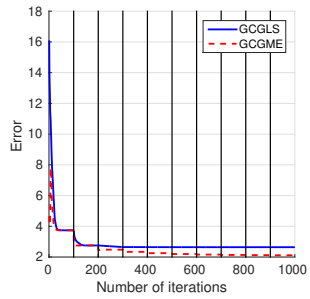
(a) GCGLS



(b) GCGME



(c) Value of the objective function

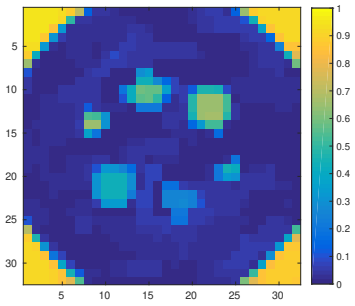


(d) Error

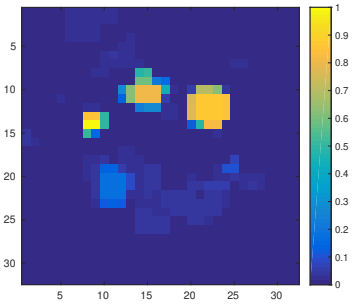
Figure 7: Results obtained using the GCGLS and GCGME algorithms when LASSO regularization is used. The regularization parameter is set to $\tau = 3 \cdot 10^{-5}$. The value of objective function (46) and the error are measured in the ℓ_2 norm. The number of CG iterations is increased to 100. In the plots, the vertical black lines indicate the start of a new fixed-point iteration.

6.2.2 Total variation

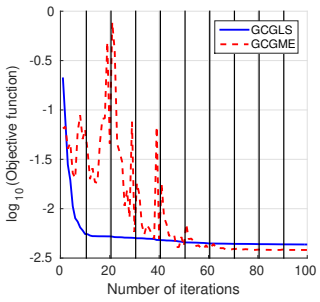
Still using the ℓ_1 penalty, we now define $\mathbf{F}$ as the total variation operator. Again, 10 IRLS iterations and 10 CG iterations per IRLS iteration are used. The regularization parameter is set to 10^{-5}. As in the ℓ_2 case, on average, GCGLS and GCGME need 0.023 and 0.024 seconds per iteration respectively. Figure 8 shows the results. The GCGME solution has a smaller error, as can be seen in figure 8d. This is consistent with the images that are obtained using the two different methods. The GCGME image resembles the original much more than the GCGLS image does.



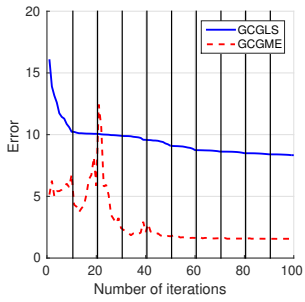
(a) GCGLS



(b) GCGME



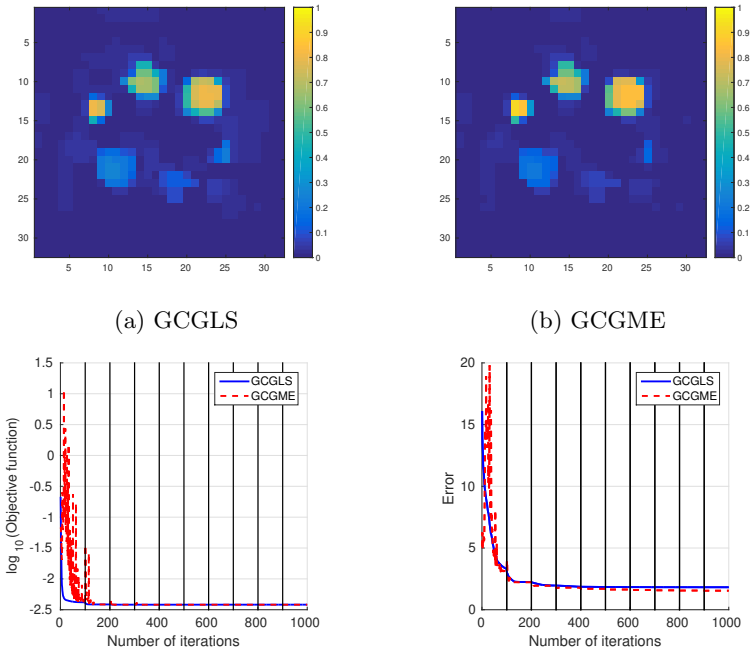
(c) Value of the objective function



(d) Error

Figure 8: Results obtained using the GCGLS and GCGME algorithms when total variation regularization is used. The regularization parameter is set to $\tau = 10^{-5}$. The value of objective function (46) and the error are measured in the ℓ_2 norm. The number of CG iterations is set to 10. In the plots, the vertical black lines indicate the start of a new fixed-point iteration.

Increasing the number of CG iterations to 100 yields the results in figure 9. Again, the GCGME image looks about the same as before, whereas the GCGLS one looks different and is now almost the same as the GCGME image.



(a) GCGLS (b) GCGME

(c) Value of the objective function (d) Error

Figure 9: Results obtained using the GCGLS and GCGME algorithms when total variation regularization is used. The regularization parameter is set to $\tau = 10^{-5}$. The value of objective function (46) and the error are measured in the ℓ_2 norm. The number of CG iterations is increased to 100. In the plots, the vertical black lines indicate the start of a new fixed-point iteration.

7 Conclusion

We considered the weighted and regularized least squares problem. A second set of normal equations was derived, which allowed us to generalize the Conjugate Gradient Minimal Error (CGME) method to include nontrivial weighting and regularization matrices.

We compared our GCGME method to the well-known GCGLS method by applying both to data simulated in a low-field MRI setting. We considered different regularization operators $\mathbf{F}$. First, we set $\mathbf{F} = \mathbf{I}$. Then, we defined $\mathbf{F}$ as a first order difference matrix that determines the size of the jumps between neighboring pixels. In both cases, the GCGME algorithm was outperformed by GCGLS; GCGLS needs a much lower number of iterations to converge.

Alternatively, we considered an ℓ_1 penalty, with the same regularization operators as before, the identity matrix and the total variation matrix. We used IRLS to solve the ℓ_1 case. GCGLS needs a high number of CG iterations to converge, while for GCGME, this number is low (10 is sufficient). It is interesting to note that when the number of CG iterations for GCGLS is set to 10 as well, GCGLS appears to have reached convergence after 10 IRLS iterations. GCGLS yields an image with artifacts in the form of three additional shapes in the lower half of the image, as well as regions of nonzeros in the corners of the

image. However, convergence is not actually attained.

While increasing the number of CG iterations leads to GCGLS images of higher quality, it actually seems to lead to images of poorer quality in the GCGME case. Increasing the number of CG iterations leads to more visible artifacts in the form of three shapes in the lower half of the image for GCGME. The reason these shapes appear is because the magnetic field is almost symmetrical, which means that it is very difficult to distinguish one frequency in the upper half of the image from its counterpart in the lower half of the image. Interestingly, GCGME seems to have a regularizing effect that prevents (to some extent) this symmetry from corrupting the image. This effect is more pronounced when the number of CG iterations is low.

GCGME outperforming GCGLS in the ℓ_1 case can be explained by the fact that as we get closer to the solution, many elements of the vector $|\mathbf{F}\mathbf{x}_k|$ will converge to zero, due to the sparsity-enforcing properties of the ℓ_1 penalty. This means that, after a few fixed-point iterations, the condition number of $\mathbf{R}_k$ will be much larger than the condition number of $\mathbf{C} = \mathbf{I}$, making GCGME the preferred algorithm.

8 Acknowledgments

This research is supported by NWO-WOTRO Grant W07.303.101. We gratefully acknowledge Peter Sonneveld for useful discussions regarding Krylov methods. We would like to thank the low-field MRI team members of the Leiden University Medical Center, the Electronic and Mechanical Support Division (DEMO) in Delft, Pennsylvania State University and Mbarara University of Science and Technology for their insight and expertise.

References

[1] M. Arioli and D. Orban. Iterative methods for symmetric quasi-definite linear systems—Part i: Theory. *Cahier du GERAD G-2013-32, GERAD, Montréal, QC, Canada*, 2013.

[2] R. Barrett, M. Berry, T. F. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. Van der Vorst. *Templates for the solution of linear systems: building blocks for iterative methods*. SIAM, 1994.

[3] Å. Björck. *Numerical methods for least squares problems*. SIAM, 1996.

[4] C. Z. Cooley, J. P. Stockmann, B. D. Armstrong, M. Sarracanie, M. H. Lev, M. S. Rosen, and L. L. Wald. Two-dimensional imaging in a lightweight portable mri scanner without gradient coils. *Magnetic Resonance in Medicine*, 73(2):872–883.

[5] E. J. Craig. The N-step Iteration Procedures. *Studies in Applied Mathematics*, 34(1-4):64–73, 1955.

[6] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375. Springer Science & Business Media, 1996.

[7] J. A. Fessler. Model-based image reconstruction for MRI. *IEEE Signal Processing Magazine*, 27(4):81–89, 2010.

[8] M. Hanke. *Conjugate gradient type methods for ill-posed problems*, volume 327. CRC Press, 1995.

[9] M. Hanke. The minimal error conjugate gradient method is a regularization method. *Proceedings of the American Mathematical Society*, 123(11):3487–3497, 1995.

[10] P. C. Hansen. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM review*, 34(4):561–580, 1992.

[11] P. C. Hansen. *Discrete inverse problems: insight and algorithms*. SIAM, 2010.

[12] M. R. Hestenes and E. Stiefel. *Methods of conjugate gradients for solving linear systems*, volume 49. NBS, 1952.

[13] J. King. A minimal error conjugate gradient method for ill-posed problems. *Journal of optimization theory and applications*, 60(2):297–304, 1989.

[14] Z. P. Liang and P. C. Lauterbur. *Principles of Magnetic Resonance Imaging: A Signal Processing Perspective*. SPIE Optical Engineering Press, 2000.

[15] A. S. Nemirovskii. The regularizing properties of the adjoint gradient method in ill-posed problems. *USSR Computational Mathematics and Mathematical Physics*, 26(2):7–16, 1986.

[16] Y. Saad. *Iterative methods for sparse linear systems*. SIAM, 2003.

[17] M. A. Saunders. Solution of sparse rectangular systems using LSQR and CRAIG. *BIT Numerical Mathematics*, 35(4):588–604, 1995.

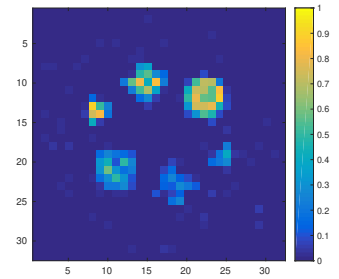
[18] B. P. Sutton, D. C. Noll, and J. A. Fessler. Fast, iterative image reconstruction for MRI in the presence of field inhomogeneities. *IEEE transactions on medical imaging*, 22(2):178–188, 2003.

[19] R. Tibshirani. Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

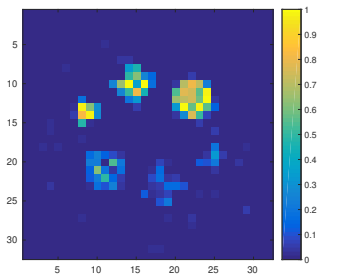
[20] A. N. Tikhonov. On the solution of ill-posed problems and the method of regularization. In *Doklady Akademii Nauk*, volume 151, pages 501–504. Russian Academy of Sciences, 1963.

[21] H. Weyl. Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung). *Mathematische Annalen*, 71(4):441–479, Dec 1912.

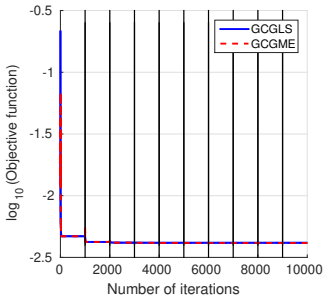
9 Appendix



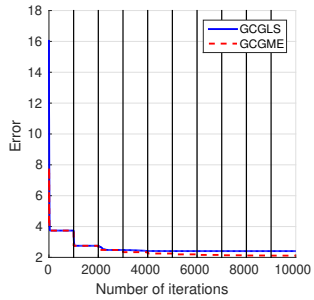
(a) GCGLS



(b) GCGME

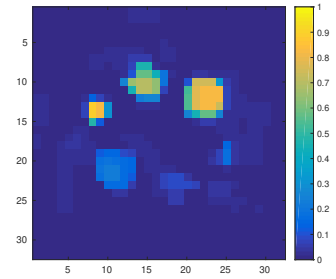


(c) Value of the objective function

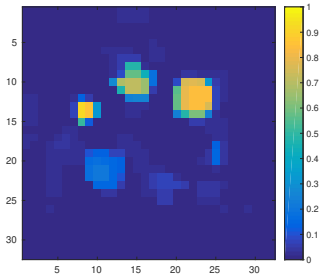


(d) Error

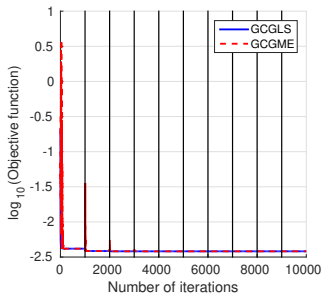
Figure 10: Results obtained using the GCGLS and GCGME algorithms when LASSO regularization is used. The regularization parameter is set to $\tau = 3 \cdot 10^{-5}$. The value of objective function (46) and the error are measured in the ℓ_2 norm. The number of CG iterations is increased to 1000. In the plots, the vertical black lines indicate the start of a new fixed-point iteration.



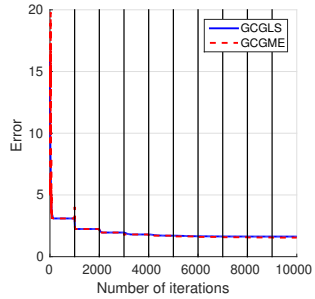
(a) GCGLS



(b) GCGME



(c) Value of the objective function



(d) Error

Figure 11: Results obtained using the GCGLS and GCGME algorithms when total variation regularization is used. The regularization parameter is set to $\tau = 10^{-5}$. The value of objective function (46) and the error are measured in the ℓ_2 norm. The number of CG iterations is increased to 1000. In the plots, the vertical black lines indicate the start of a new fixed-point iteration.