CG variants 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

Abstract In an earlier paper, we generalized the CGME (Conjugate Gradient Minimal Error) algorithm to the ℓ_2-regularized weighted least-squares problem. Here, we use this Generalized CGME method to reconstruct images from actual signals measured using a low-field MRI scanner. We analyze the convergence of both GCGME and the classical Generalized Conjugate Gradient Least Squares (GCGLS) method for the simple case when a Laplace operator is used as a regularizer and indicate when GCGME is to be preferred in terms of convergence speed. We also consider a more complicated ℓ_1-penalty in a compressed sensing framework.

1 Introduction

In Magnetic Resonance Imaging (MRI), the measured signal $\mathbf{b}$ is related to $\mathbf{x}$, the object being imaged, by a Fourier Transform:

$$\mathbf{b} = \mathcal{F}\mathbf{x} + \mathbf{v}. \tag{1}$$

Here, $\mathbf{v}$ denotes a noise vector. Based on measurements $\mathbf{b}$, we will reconstruct $\mathbf{x}$, which makes this an inverse problem. In this work, we will assume the object of interest to be 2D, which means that $\mathcal{F}$ is a 2D Fourier Transform operator. However, all the results can be extended to 3D.

In conventional MRI, the signal-to-noise ratio (SNR) is so high that applying an Inverse Fourier Transform usually results in an image of very good quality. This

M.L. de Leeuw den Bouter
Delft Institute of Applied Mathematics, Delft University of Technology, Van Mourik Broekmanweg 6, 2628 XE Delft, The Netherlands, e-mail: M.L.deLeeuwdenBouter-1@tudelft.nl

M.B. van Gijzen
Delft Institute of Applied Mathematics, Delft University of Technology, Van Mourik Broekmanweg 6, 2628 XE Delft, The Netherlands, e-mail: M.B.vanGijzen@tudelft.nl

R.F. Remis
Circuits and Systems, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands, e-mail: R.F.Remis@tudelft.nl

is because superconducting magnets are used to generate strong magnetic fields strengths of several tesla and the SNR is higher in the case of a stronger magnetic field. In [8], O'Reilly et al. describe a low-field MRI scanner based on a configuration of permanent magnets. The magnetic field strength inside this scanner is 50 mT, whereas conventional scanners have background fields of several teslas. For very noisy signals, it can be useful to minimize a regularized least-squares problem of the form

$$\arg\min_{\mathbf{x}} \frac{1}{2}||\mathbf{b} - \mathcal{F}\mathbf{x}||^2_{\mathbf{C}^{-1}} + \frac{1}{2}\tau||\mathbf{x}||^2_{\mathbf{R}}, \tag{2}$$

instead of disregarding the noise $\mathbf{v}$ and solving equation (1) for $\mathbf{x}$. In equation (2), the regularization parameter τ determines the tradeoff between the least-squares term $||\mathbf{b} - \mathcal{F}\mathbf{x}||^2_{\mathbf{C}^{-1}}$ and the regularization term $||\mathbf{x}||^2_{\mathbf{R}}$. In the least-squares term, $\mathbf{C}$ denotes the covariance matrix of the noise, and in the regularization term, $\mathbf{R}$ is a regularizing matrix, which we will assume to be Hermitian positive definite (HPD). Regularization allows us to enforce prior information we have about the solution. For a thorough exploration of the regularization of inverse problems, the reader is referred to [5].

2 GCGLS and GCGME

In [2], we introduced the Generalized Conjugate Gradient Minimal Error (GCGME) method for general form regularization. In this section we will review the main ideas. We are interested in solving minimization problems of the form

$$\arg\min_{\mathbf{x}} \frac{1}{2}||\mathbf{b} - \mathbf{A}\mathbf{x}||^2_{\mathbf{C}^{-1}} + \frac{1}{2}\tau||\mathbf{x}||^2_{\mathbf{R}}. \tag{3}$$

Note that equation (3) is of the same form as equation (2), but we have replaced $\mathcal{F}$ by a general forward model matrix $\mathbf{A}$.

Usually, minimization problem (3) is solved using the Generalized Conjugate Gradient Least-Squares (GCGLS) method. (We add the word "generalized" because CGLS is often used to denote the CG variant that solves the normal equations $\mathbf{A}^*\mathbf{A}\mathbf{x} = \mathbf{A}^*\mathbf{b}$ of the minimized least-squares problem without regularization.) By taking the gradient of equation (3) and setting it equal to zero, we find

$$\left(\mathbf{A}^*\mathbf{C}^{-1}\mathbf{A} + \tau\mathbf{R}\right)\mathbf{x} = \mathbf{A}^*\mathbf{C}^{-1}\mathbf{b}. \tag{4}$$

Equation (4) can be solved using the conjugate gradient (CG) method. Some adjustments can be made to improve stability, see for example [1], leading to the GCGLS method. By rewriting equation (3) as a constrained minimization problem, we can find another set of equations that can be used to find the solution $\mathbf{x}$. We define $\mathbf{r} = \mathbf{C}^{-1}(\mathbf{b} - \mathbf{A}\mathbf{x})$ and rewrite minimization problem (3):

$$\begin{aligned} &\min_{\mathbf{r},\mathbf{x}} \frac{1}{2}||\mathbf{r}||^2_{\mathbf{C}} + \frac{1}{2}\tau||\mathbf{x}||^2_{\mathbf{R}} \\ &\text{s.t. } \mathbf{r} = \mathbf{C}^{-1}(\mathbf{b} - \mathbf{A}\mathbf{x}) \end{aligned} \tag{5}$$

We will assume $\tau > 0$. By applying the method of Lagrange multipliers and eliminating $\mathbf{x}$, we get

$$\left(\frac{1}{\tau}\mathbf{A}\mathbf{R}^{-1}\mathbf{A}^{*} + \mathbf{C}\right)\mathbf{r} = \mathbf{b}. \tag{6}$$

Additionally, the following relationship between $\mathbf{r}$ and $\mathbf{x}$ holds:

$$\mathbf{x} = \frac{1}{\tau}\mathbf{R}^{-1}\mathbf{A}^{*}\mathbf{r}. \tag{7}$$

So by applying CG to equation (6) and subsequently solving equation (7) for $\mathbf{x}$, we can obtain our solution. The resulting algorithm, which we call Generalized Conjugate Gradient Minimal Error (GCGME), is given below.

Algorithm 1 GCGME

Require: $\mathbf{A} \in \mathbb{C}^{M\times N}$, $\mathbf{C} \in \mathbb{C}^{M\times M}$, $\mathbf{R} \in \mathbb{C}^{N\times N}$, $\mathbf{r}_0 \in \mathbb{C}^{M}$, $\mathbf{b} \in \mathbb{C}^{M}$, $\tau \in \mathbb{R}_{>0}$;
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

2.1 Comparison of the condition numbers: a simple case

In this section we consider a very simple but illustrative case that allows us to analyze the condition numbers, and hence the convergence speed, of GCGME and GCGLS. We demonstrate, depending on the regularization parameter, which method is to be preferred. We set $\mathbf{A} = \mathcal{F}$ and the noise is assumed to be white noise, so $\mathbf{C} = \mathbf{I}$. We define the regularization matrix to be the discretized 2D Laplacian $\mathbf{L}$ complemented with Dirichlet boundary conditions. Choosing the regularization matrix in this way means that large jumps in the reconstructed image $\mathbf{x}$ are discouraged. In that case, GCGLS solves

$$(\mathbf{I} + \tau\mathbf{L})\,\mathbf{x} = \mathcal{F}^{*}\mathbf{b}, \tag{8}$$

where $\mathcal{F}^{*} = \mathcal{F}^{-1}$ is the inverse 2D Fourier Transform. GCGME solves

$$\left(\frac{1}{\tau}\boldsymbol{\mathcal{F}}\mathbf{L}^{-1}\boldsymbol{\mathcal{F}}^{*}+\mathbf{I}\right)\mathbf{r}=\mathbf{b}\,, \tag{9}$$

$$\mathbf{x}=\frac{1}{\tau}\mathbf{L}^{-1}\boldsymbol{\mathcal{F}}^{*}\mathbf{r}.$$

The convergence speed of GCGLS and GCGME depends on the condition number of the matrices $\mathbf{I}+\tau\mathbf{L}$ and $\frac{1}{\tau}\boldsymbol{\mathcal{F}}\mathbf{L}^{-1}\boldsymbol{\mathcal{F}}^{*}+\mathbf{I}$, respectively. The eigenvalues of the Laplacian $\mathbf{L}$ are well-known and hence we can find explicit expressions for the condition numbers. For GCGLS, we have

$$\kappa_2\left(\mathbf{I}+\tau\mathbf{L}\right)=\frac{1+8\tau\cos^2\left(\frac{\pi}{2}\frac{1}{N+1}\right)}{1+8\tau\sin^2\left(\frac{\pi}{2}\frac{1}{N+1}\right)}. \tag{10}$$

Here, we assume that our image consists of $N\times N$ pixels. For GCGME, we make use of the fact that $\boldsymbol{\mathcal{F}}\mathbf{L}^{-1}\boldsymbol{\mathcal{F}}^{*}$ is a similarity transformation and therefore has the same eigenvalues as $\mathbf{L}^{-1}$, yielding

$$\kappa_2\left(\frac{1}{\tau}\boldsymbol{\mathcal{F}}\mathbf{L}^{-1}\boldsymbol{\mathcal{F}}^{*}+\mathbf{I}\right)=\frac{1+\frac{1}{8\tau\sin^2\left(\frac{\pi}{2}\frac{1}{N+1}\right)}}{1+\frac{1}{8\tau\cos^2\left(\frac{\pi}{2}\frac{1}{N+1}\right)}}. \tag{11}$$

These condition numbers can be shown to be equal when

$$\tau^{*}=\frac{1}{8\cos\left(\frac{\pi}{2(N+1)}\right)\sin\left(\frac{\pi}{2(N+1)}\right)}. \tag{12}$$

Figure 1 shows a plot of the condition numbers as a function of the value of the regularization parameter τ, in case $N=128$. We observe that when $\tau<\tau^{*}$, GCGLS has a smaller condition number, whereas GCGME has a smaller condition number when $\tau>\tau^{*}$ is large. Therefore, we expect GCGME to attain faster convergence for large τ.

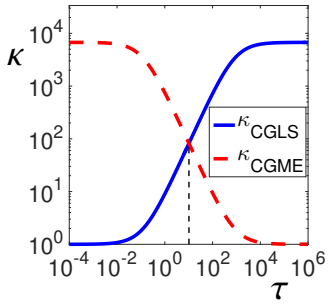


Fig. 1: Condition numbers of the GCGLS matrix $\mathbf{I}+\tau\mathbf{L}$ and the GCGME matrix $\frac{1}{\tau}\boldsymbol{\mathcal{F}}\mathbf{L}^{-1}\boldsymbol{\mathcal{F}}^{*}+\mathbf{I}$ as a function of the value of the regularization parameter τ.

2.2 GCGLS and GCGME for IRLS

The ℓ_2-penalty tends to lead to overly blurry images, due to the quadratic penalty term. Therefore, we are more interested in the ℓ_p-regularized least squares problem with $p \in (0, 1]$:

$$\min_{\mathbf{x}} \frac{1}{2}||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2 + \frac{1}{p}\tau||\mathbf{F}\mathbf{x}||_p^p, \tag{13}$$

For the ℓ_p-penalty with $p \in (0, 1]$, the blurring effect is less pronounced. Additionally, the ℓ_p-penalty induces sparsity in $\mathbf{F}\mathbf{x}$, see for example [4]. However, solving minimization problem (13) is not as straightforward as equation (3). One way of solving it is by using Iterative Reweighted Least Squares (IRLS). This means that we replace minimization problem (13) by a sequence of ℓ_2-regularized problems of the same form as equation (3). Given an estimate $\mathbf{x}_k$ of the solution $\mathbf{x}$, the matrix $\mathbf{R}_k$ in the penalty term is recalculated based on $\mathbf{x}_k$:

$$\mathbf{R}_k = \mathbf{F}^*\mathbf{D}_k\mathbf{F}, \qquad \mathbf{D}_k = \text{diag}\left(\frac{1}{|\mathbf{F}\mathbf{x}_k|^{2-p}}\right). \tag{14}$$

So in each IRLS step, one minimization problem of the form (3) is solved. We will compare GCGLS and GCGME for this step. In case $\mathbf{F}$ is an invertible matrix and $\mathbf{R}_k^{-1} = \mathbf{F}^{-1}\mathbf{D}_k^{-1}(\mathbf{F}^H)^{-1}$, with

$$\mathbf{D}_k^{-1} = \text{diag}\left(|\mathbf{F}\mathbf{x}_k|^{2-p}\right). \tag{15}$$

When GCGME is used, we can take advantage of this structure, instead of calculating $\mathbf{R}_k$ and working with its inverse. Moreover, when $\mathbf{F}$ is an orthogonal matrix, no additional computations are necessary to compute inverses.

In [2], we showed that when

$$\kappa_2(\mathbf{R}) \gg \kappa_2(\mathbf{C}), \tag{16}$$

GCGME is expected to exhibit faster converge than GCGLS. When the sparsifying ℓ_p-penalty with $p \in (0, 1]$ is used, some elements of D_k will tend to infinity. Therefore, $\mathbf{R}$ is expected to become increasingly ill-conditioned, in which case Equation (16) holds. Therefore, we expect GCGLS to be outperformed by GCGME in terms of convergence speed.

3 Experiments

Experiments were carried out using the low-field MRI scanner described in [8], a picture of which is shown in Figure 2a. Inside the scanner, the magnetic field generated by the configuration of magnets is approximately homogeneous. Linear gradient fields are applied before and during the pulse for phase and frequency encoding. These steps ensure that the resulting signal is essentially equal to the Fourier Transform of the object inside the scanner. For an introduction to the principles of MRI, the reader is referred to [6]. The object being imaged, see Figure 2b, is a real-life version of the Shepp-Logan phantom, which was introduced in [9]. It is

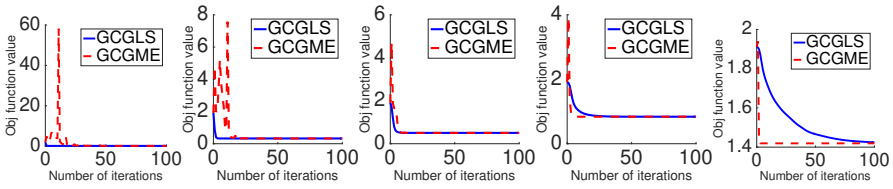
approximately 10 cm in diameter. This phantom is often used to test reconstruction algorithms used in tomography. The sampling rate was set to 20 μs, corresponding to a bandwidth of 50 kHz. A spin echo pulsing sequence was used with an echo time T_E of 10 ms and a repetition time T_R of 500 ms. The length of the RF pulse was 100 μs. The Field of View (FoV) was 12x12 cm^2, with the target image having 128x128 pixels. No slice selection was carried out.

(a) Low-field MRI scanner (b) Object being imaged

Fig. 2: Experimental setup

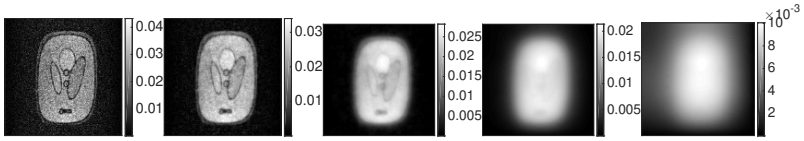
4 Numerical results

First, we solve minimization problem (3) with $\mathbf{A} = \mathcal{F}$, $\mathbf{C} = \mathbf{I}$ and $\mathbf{R} = \mathbf{L}$, which is the scenario we reviewed earlier. GCGLS and GCGME solve different normal equations, so a comparison using a stopping criterion based on residuals would not be fair. Instead, we use a fixed number of CG iterations for both methods. For the ℓ_2 case, we use 100 iterations. Figure 3 shows plots of the value of objective function (2) with $\mathbf{R} = \mathbf{L}$ as a function of the iteration number for 5 different values of the regularization parameter τ. We observe that in all cases, both methods lead to the same objective function value, as expected. For $\tau = 10$, which is approximately equal to τ^*, we note that both methods converge equally fast. For smaller values of τ, GCGLS converges faster while for larger values, GCGME shows faster convergence. The corresponding images are shown in Figure 4. Both methods need the same amount of time per iteration.



(a) $\tau = 0.1$ (b) $\tau = 1$ (c) $\tau = 10$ (d) $\tau = 100$ (e) $\tau = 1000$

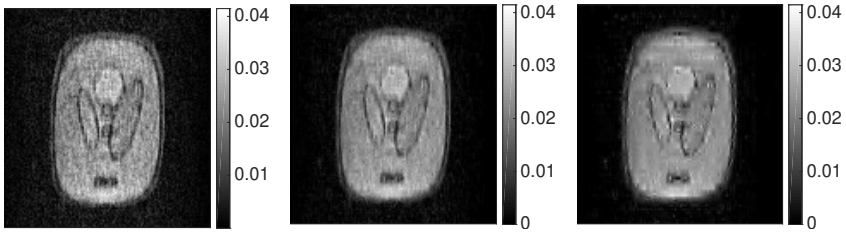
Fig. 3: Objective function value as a function of the iteration number for different values of the regularization parameter τ.



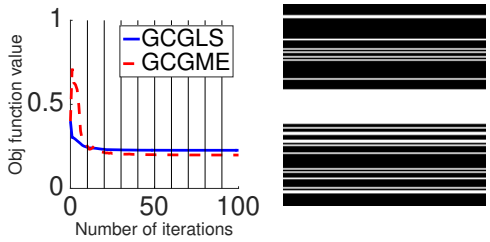
(a) $\tau = 0.1$ (b) $\tau = 1$ (c) $\tau = 10$ (d) $\tau = 100$ (e) $\tau = 1000$

Fig. 4: Reconstructed images for different values of the regularization parameter τ.

In MRI, scan times tend to be long. They can be reduced by using compressed sensing. In compressed sensing, the number of data points acquired is reduced, compared to traditional scans. This can be done by measuring a subset of the lines in k-space, or the frequency domain. For more information about compressed sensing in MRI, [7] can be consulted. We will use the notation $\mathcal{F}_u$ to denote the Fourier Transform of the undersampled measurements. One of the assumptions made in compressed sensing is that the image is sparse in some known transform domain, for example a wavelet transform. We also investigate the two CG variants in a compressed sensing framework with an undersampling factor of 3.



(a) No regularization (b) GCGLS result (c) GCGME result



(d) Objective function (e) Lines in k-space

Fig. 5: Reconstructed images using (a) only the inverse Fourier Transform, (b) and (c) the two different CG variants. (d) shows a plot of the objective function value as a function of the iteration number for both methods. The vertical black lines indicate the start of a new IRLS iteration. (e) shows the lines in k-space that were used for reconstruction.

We solve minimization problem (13) with $\mathbf{A} = \mathcal{F}_u$, $\mathbf{C} = \mathbf{I}$, $\mathbf{F} = \mathbf{W}$, $\tau = 6 \cdot 10^{-3}$ and $p = 1$. The regularization parameter is chosen heuristically. Here, $\mathbf{W}$ is the 2D Daubechies wavelet transform [3]. We choose $\mathcal{F}_u\mathbf{b}$, which is shown in Figure 5a, as our initial guess. We use 10 IRLS iterations and in each of these, 10 CG iterations are carried out. Figure 5 shows the reconstructed images and the value of the objective function as a function of the iteration number. GCGME shows rapid convergence, whereas the convergence of GCGLS is so slow that it seems that GCGLS has converged to a higher objective function value than GCGME. However, both methods converge to the same value if the number of GCGLS iterations is increased significantly, see [2]. GCGLS and GCGME need the same amount of time per iteration.

5 Conclusion

We analyzed the condition numbers of the matrices used in GCGME and GCGLS in the simple but illustrative case where the discretized Laplacian is used as the regularization matrix. The value of the regularization parameter τ^* determines which method is to be preferred in terms of convergence speed. We can easily calculate τ^*, the value for which both methods have the same condition number. For $\tau < \tau^*$, GCGLS is expected to converge faster and for $\tau > \tau^*$, GCGME is to be preferred. We applied both methods to data measured using a low-field MRI scanner and our numerical results show that the two methods behave as expected.

We also considered the more relevant case of an ℓ_1-regularization penalty in a compressed sensing framework and used IRLS to solve this problem. Inside each IRLS iteration, GCGLS or GCGME can be used as a building block. Due to the sparsifying properties of the ℓ_p-penalty with $p \in (0, 1]$, the reweighting of the regularization matrix leads to an increasingly ill-conditioned matrix, which corresponds to the regime in which GCGME is expected to show rapid convergence. Our numerical results show that indeed, GCGME converges much faster than GCGLS for this problem.

References

[1] Björck, Å.: Numerical methods for least squares problems. SIAM (1996)
[2] de Leeuw den Bouter, M.L., van Gijzen, M.B., Remis, R.F.: Conjugate gradient variants for ℓ_p-regularized image reconstruction in low-field MRI. SN Applied Sciences **1**(12), 1736 (2019)
[3] Daubechies, I.: Ten lectures on wavelets, vol. 61. Siam (1992)
[4] Elad, M.: Sparse and redundant representations: from theory to applications in signal and image processing. Springer Science & Business Media (2010)
[5] Engl, H.W., Hanke, M., Neubauer, A.: Regularization of inverse problems, vol. 375. Springer Science & Business Media (1996)
[6] Liang, Z.P., Lauterbur, P.C.: Principles of Magnetic Resonance Imaging: A Signal Processing Perspective. SPIE Optical Engineering Press (2000)
[7] Lustig, M., Donoho, D.L., Santos, J.M., Pauly, J.M.: Compressed sensing MRI. IEEE signal processing magazine **25**(2), 72 (2008)
[8] O'Reilly, T., Teeuwisse, W., Webb, A.: Three-dimensional MRI in a homogenous 27 cm diameter bore halbach array magnet. Journal of Magnetic Resonance **307**, 106578 (2019)
[9] Shepp, L.A., Logan, B.F.: The Fourier reconstruction of a head section. IEEE Transactions on Nuclear Science **21**(3), 21–43 (1974). DOI 10.1109/TNS.1974.6499235