AMSC 664 Final Presentation

Introduction of Bias into Magnetic Resonance Relaxometry Modeling to Break the CRLB Barrier in Parameter Estimation

Zezheng Song [1]

Advisor: Dr. Richard Spencer [2]

[1]AMSC Program, Department of Mathematics, University of Maryland, College Park, MD

[2]National Institute on Aging, National Institutes of Health, Baltimore, MD

2021/05/06

Presentation Outline

Background

Process of Magnetic Resonance Imaging (MRI)

- Signal generation based on magnetic resonance properties of material or tissue
- Relaxation processes
- Signal detection

Background

- In biomedical Magnetic Resonance Imaging (MRI), an important task is to recover parameter values for underlying MRI models, which could provide important biophysical information.
- The most common multi-component relaxometry model is the bi-exponential model, which describes the simultaneous transverse relaxation, or signal loss, of two tissue components that co-exist within a sample or tissue, i.e.,

$$S(\mathbf{TE}; c_1, c_2, T_{21}, T_{22}) = c_1 \exp\left(-\mathbf{TE}/T_{21}\right) + c_2 \exp\left(-\mathbf{TE}/T_{22}\right)$$

  where $\mathbf{TE}$ is the echo time, with the signal sampling times then given by $\{kTE\}_{k=1}^{n}$, T_{21} and T_{22} are transverse relaxation time constants, and c_1, c_2 are component fractions.
- The bi-exponential model is used for assessment of cartilage degeneration in osteoarthritis:
  - rapidly relaxing component: proteoglycan
  - slowly relaxing component: less-bound water

Motivation

- We are currently assuming known values for c_1 and c_2 and studying the tissue characteristics by estimating the two relaxation times, T_{21} and T_{22}, which relate to tissue hydration and microscopic organization.
- We seek to develop methods to decrease the Mean Squared Error (MSE) of parameter estimation below the conventional lower limit of the Cramér-Rao lower bound. In addition, we seek to reduce MSE below the values obtained through conventional non-linear least squares.

Definition

In the field of statistics, the Cramér–Rao lower bound (CRLB) gives a theoretical lower bound on the variance of an unbiased estimator of model parameters.

- Consider the the parameter vector $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_d]^T \in \mathbb{R}^d$, with probability density function $f(x; \boldsymbol{\theta})$, and the Fisher information matrix is a $d \times d$ matrix given by $I_{m,k} = -\mathrm{E}\left[\frac{\partial^2}{\partial\theta_m \partial\theta_k} \log f(x; \boldsymbol{\theta})\right]$. Let $\boldsymbol{T}(X)$ be an estimator, and denote its expectation vector $\mathrm{E}[\boldsymbol{T}(X)]$ by $\boldsymbol{\psi}(\boldsymbol{\theta})$, then the Cramér–Rao bound states that the covariance matrix of $\boldsymbol{T}(X)$ satisfies

$$\mathrm{cov}_{\boldsymbol{\theta}}(\boldsymbol{T}(X)) \geq \frac{\partial \boldsymbol{\psi}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} [I(\boldsymbol{\theta})]^{-1} \left(\frac{\partial \boldsymbol{\psi}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T \tag{1.1}$$

Previous Studies

- Previous work has focused on providing CRLB, which gives the lower bound on the variance of unbiased estimators of parameters in MRI models.
- CRLB gives the best we can do for Mean Squared Error (MSE) of an unbiased estimator, and we wish to minimize MSE below CRLB. To do this, we artificially introduce bias through regularization. With this expanded space, we seek an estimator that provides MSE less than CRLB.
- Additionally, we seek an estimator that improves upon conventional Nonlinear Least Squares (NLLS). Note that since NLLS is biased in the presence of noise, CRLB does not strictly apply. However, we will show that with reasonable SNR, the CRLB is very close to the MSE for NLLS.

Previous Studies

- Eldar, et al. (2006) [Eld06] showed that the performance of estimators could be improved, provided that we introduce some bias.
- Box, et al. (1971) [Box71] did important early first order approximation analysis of bias in nonlinear estimation, using Taylor expansion.

Our approach: Introduction of Regularization

- Our metric for the performance on an estimator is the MSE, defined as

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{E}_\theta\left[(\hat{\theta} - \theta)^2\right]$$

  In addition, $\mathrm{MSE}(\hat{\theta}) = \mathrm{Var}(\hat{\theta}) + \mathrm{Bias}(\hat{\theta}, \theta)^2$, where $\hat{\theta}$ is an estimator with respect to an unknown parameter θ.
- We propose to introduce regularization to the estimation procedure, which could adjust to a better MSE.
- How to select the appropriate regularization parameter λ?

Mathematical Formulation

We formulate the following general regularized nonlinear least squares problem:

$$\hat{\mathbf{p}}_\lambda = \underset{\mathbf{p}}{\operatorname{argmin}} \left\{ \|\mathbf{G}(\mathbf{p}) - \mathbf{d}\|_2^2 + \lambda^2 \|\mathbf{L}\mathbf{p}\|_2^2 \right\} \tag{1.2}$$

- $\mathbf{d}$: Noisy data vector.
- $\mathbf{p}$: Underlying parameters.
- $\mathbf{L}$: Weighting matrix L.
- $\mathbf{G}$: Physical model of the experiment, depending on particular models.
- λ: Regularization parameter.

First, consider the simplified version of

$$\hat{\mathbf{p}}_\lambda = \underset{\mathbf{p}}{\operatorname{argmin}} \left\{ \|\mathbf{G}(\mathbf{p}) - \mathbf{d}\|_2^2 + \lambda^2 \|\mathbf{p}\|_2^2 \right\} \quad (1.3)$$

and assume $\mathbf{G}$ is a matrix of appropriate dimension, then we have

$$\hat{\mathbf{p}}_\lambda = \underset{\mathbf{p}}{\operatorname{argmin}} \left\{ \|\mathbf{G}\mathbf{p} - \mathbf{d}\|_2^2 + \lambda^2 \|\mathbf{p}\|_2^2 \right\} \quad (1.4)$$

which has a closed-form solution $\hat{\mathbf{p}}_\lambda = \left(\mathbf{G}^{\mathbf{T}}\mathbf{G} + \lambda^{\mathbf{2}}\mathbf{I}\right)^{-\mathbf{1}} \mathbf{G}^{\mathbf{T}}\mathbf{d}$, where $\mathbf{I}$ is the identity matrix.

- The role of λ is to perturb the ill-conditioned matrix $\mathbf{G}^{\mathbf{T}}\mathbf{G}$, which is precisely what is required to improve its condition number.

Spencer and Bi (2020) derived an explicit expression for the covariance matrix of the solution to the optimization problem (1.2):

$$\hat{\mathbf{p}}_\lambda = \underset{\mathbf{p}}{\operatorname{argmin}} \left\{ \|\mathbf{G}(\mathbf{p}) - \mathbf{d}\|_2^2 + \lambda^2 \|\mathbf{L}\mathbf{p}\|_2^2 \right\}$$

given by

$$\operatorname{Cov}(\mathbf{p}_\lambda^*) \approx \sigma_\varepsilon^2 \left(\mathbf{J}^T\mathbf{J} + \lambda^2\mathbf{L}^T\mathbf{L}\right)^{-1} \mathbf{J}^T\mathbf{J} \left(\left(\mathbf{J}^T\mathbf{J} + \lambda^2\mathbf{L}^T\mathbf{L}\right)^{-1}\right)^T \tag{1.5}$$

where σ_ϵ is the standard deviation of noise of data, $\mathbf{L}$ is the weighting matrix, λ is the regularization parameter, and $\mathbf{J}$ is the Jacobian of $\mathbf{G}$.

- If comparing with the non-regularized covariance matrix, the elements of the covariance matrix are greatly reduced, indicating the improved stability of regularized solution.

Mathematical Formulation

To solve nonlinear least squares problem (1.2),

$$\hat{\mathbf{p}}_\lambda = \underset{\mathbf{p}}{\operatorname{argmin}} \left\{ \|\mathbf{G}(\mathbf{p}) - \mathbf{d}\|_2^2 + \lambda^2 \|\mathbf{L}\mathbf{p}\|_2^2 \right\}$$

we implemented various optimization methods:

- Grid Search + Mesh Adaptive Direct Search
- Gauss Newton
- VARPRO [OR13]
- Levenberg-Marquardt Method

Project Goals

- Evaluate a suitable degree of regularization, through the value of λ, that we should introduce to achieve MSE less than CRLB.
- If such λ exists, evaluate its robustness with respect to range of possible parameter values. In practical applications, we will assume the realistic case of having some a priori knowledge of plausible parameter ranges.

Mono-exponential Model

To establish methodology with a simpler but still-relevant problem, we first analyze the simpler but still-important mono-exponential model:

$$S(\mathbf{TE}; c, T) = c \exp\left(-\mathbf{TE}/T\right)$$

where $\mathbf{TE} = \{TE_k\}_{k=1}^{n}$ is the echo time, and c, T are parameters to be estimated. In the actual signal measured, we have additive white Gaussian noise, namely,

$$S\big(\mathbf{TE}; c, T\big) = c \exp\left(-\mathbf{TE}/T\right) + \nu \tag{1.6}$$

where ν is the noise with normal distribution, i.e., $\nu \sim \mathcal{N}(0, \sigma^2)$.

Mono-exponential Model

Now, we set up the optimization problem as follows

$$\begin{aligned} \underset{c,T}{\operatorname{argmin}} \quad & \left\| c \exp(-\mathbf{TE}/T) - \mathbf{d} \right\|_2^2 \\ \text{s.t.} \quad & c \geq 0, T \geq 0 \end{aligned} \tag{2.1}$$

where $\mathbf{TE} = \{TE_k\}_{k=1}^n$ is the echo time, $\mathbf{d} = \{d_k\}_{k=1}^n$ is the signal data of one noise realization, i.e., d_k is the datum we collect at time TE_k, for $k = 1, 2, \ldots, n$.

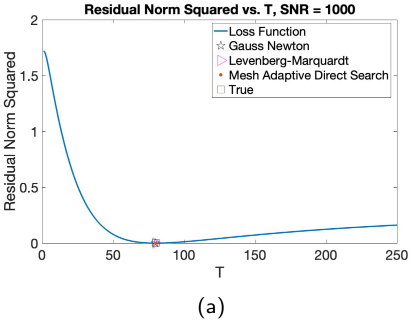


Figure: Loss function vs. T, where $c = 0.7, \mathbf{TE} = 8:8:128, SNR = 1000$.

Next we set up the regularized optimization problem as follows

$$\begin{aligned} \underset{c,T}{\operatorname{argmin}} \quad & \left\| c \exp(-\mathbf{TE}/T) - d \right\|_2^2 + \lambda^2 \left\| (c, T)^\top \right\|_2^2 \\ \text{s.t.} \quad & c \geq 0, T \geq 0 \end{aligned} \tag{2.2}$$

where $\mathbf{TE}$ is the echo time, $\mathbf{d}$ is the signal data, λ is the regularization parameter.

Algorithm

Algorithm 1: Monte Carlo Simulation

Input: Sequence of regularization parameters $\{\lambda_i\}_{i=1}^M$, number of simulations N, dimension of parameter p, weighting matrix $\mathbf{L}$, MRI model $\mathbf{G}$;

Output: Estimators of parameters sol_vec;

Data: MRI model measurements $\mathbf{d}$;

for $i = 1$ **to** M **do**

  for $j = 1$ **to** N **do**

    Using optimization solvers to solve

    $$\hat{\mathbf{p}}_\lambda = \underset{\mathbf{p}}{\operatorname{argmin}} \left\{ \|\mathbf{G}(\mathbf{p}) - \mathbf{d}(:, j)\|_2^2 + \lambda_i^2 \|\mathbf{L}\mathbf{p}\|_2^2 \right\};$$

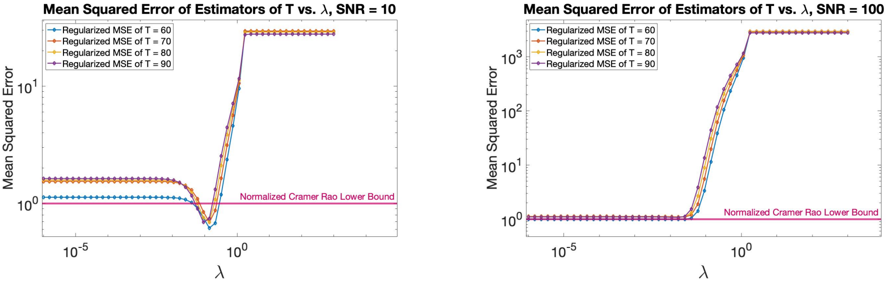
    $\text{sol\_vec}_i(j, :) = (\hat{\mathbf{p}}_\lambda)^\top$;

  end

end

Mono-exponential Model

Mean Squared Error (MSE) vs. Regularization Parameter λ



(a) $SNR = 10$ (b) $SNR = 100$

Figure: Mean Squared Error of Estimators of T vs. λ by using Levenberg-Marquardt, with different SNR.

Guarantee Analysis

- Let X denote the random variable $(\hat{\theta} - \theta)^2$, where $\hat{\theta}$ is the estimator of θ, and θ denotes c or T in the mono-exponential model.
- Using the Generalized Finite-Sample Chebyshev's inequality by Kaban, [Kab12],

$$P(|X - m| \geq ks) \leq \frac{1}{N+1} \left\lfloor \frac{N+1}{N} \left(\frac{N-1}{k^2} + 1 \right) \right\rfloor$$

where X is the random variable $(\hat{\theta} - \theta)^2$, m is the sample mean, s is the sample standard deviation, k is an arbitrary constant, and N is the number of times we sample X.

Bi-exponential Model

Bi-exponential analysis now proceeds analogously. We consider the following model

$$S(\mathbf{TE}; c_1, c_2, T_{21}, T_{22}) = c_1 \exp(-\mathbf{TE}/T_{21}) + c_2 \exp(-\mathbf{TE}/T_{22})$$

where $\mathbf{TE}$ is the echo time, and c_1, c_2, T_{21}, T_{22} are parameters to be estimated. In spectroscopy, which we consider here, the noise model is additive white Gaussian noise, namely,

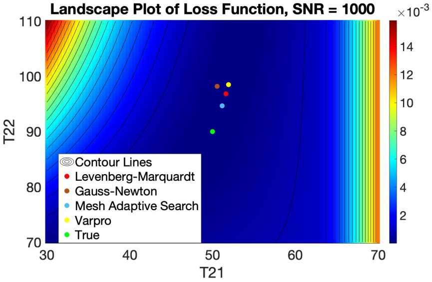
$$S(\mathbf{TE}; c_1, c_2, T_{21}, T_{22}) = c_1 \exp(-\mathbf{TE}/T_{21}) + c_2 \exp(-\mathbf{TE}/T_{22}) + \nu \qquad (3.1)$$

where ν is the noise with the normal distribution $\nu \sim \mathcal{N}(0, \sigma^2)$.

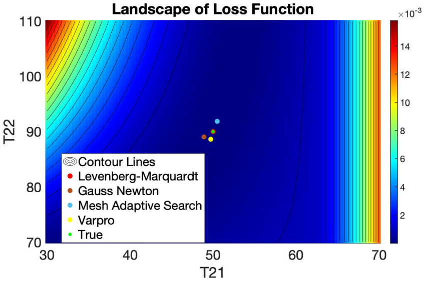
Now, we set up the optimization problem as follows

$$\begin{aligned}\underset{c_1,c_2,T_{21},T_{22}}{\operatorname{argmin}} \quad & \left\| c_1 \exp\left(-\mathbf{TE}/T_{21}\right) + c_2 \exp\left(-\mathbf{TE}/T_{22}\right) - \mathbf{d} \right\|_2^2 \\ \text{s.t.} \quad & c_1 \geq 0 \\ & c_2 \geq 0 \\ & T_{21} \geq 0 \\ & T_{22} \geq 0\end{aligned} \tag{3.2}$$

where $\mathbf{TE}$ is the echo time, $\mathbf{d}$ is the signal data of one noise realization.



(a) Data with $SNR = 1000$

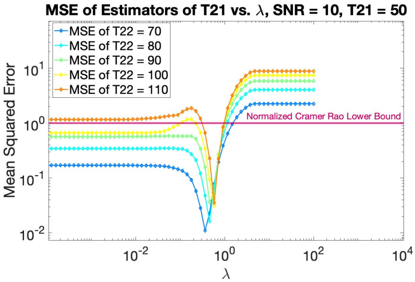


(b) Noiseless Data

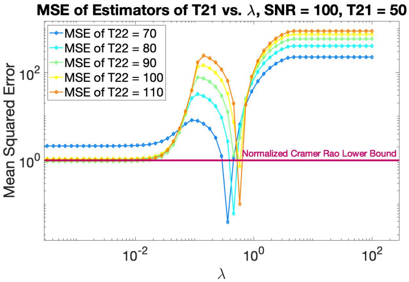
Figure: Landscape of Loss Function and Estimators

Bi-exponential Model

Mean Squared Error (MSE) vs. Regularization Parameter λ

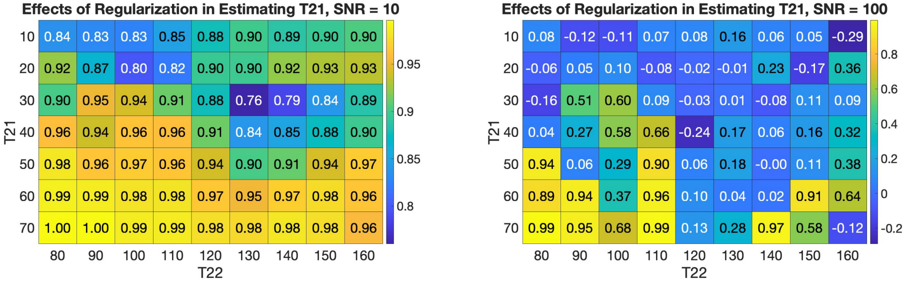


(a) $SNR = 10$



(b) $SNR = 100$

Figure: Mean Squared Error of Estimators of T_{22} vs. λ by Levenberg-Marquardt, with different SNR, $c_1 = c_2 = 0.5, T_{21} = 50$.

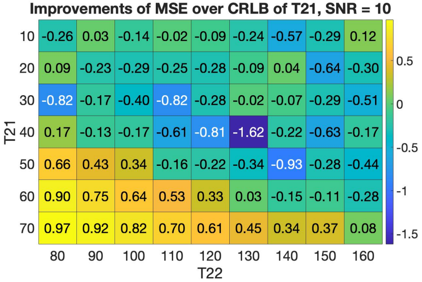


(a) $SNR = 10$ (b) $SNR = 100$

Figure: Heat map of effects of regularization in estimating T_{21} for different SNR, where $c_1 = 0.5, c_2 = 0.5$. Each entry in the table denotes improvement of MSE over CRLB, i.e., $\frac{CRLB-MSE}{CRLB}$.

Bi-exponential Model

- In practice, we are only given a prior range of possible values of parameters.
- Therefore, it is important to find optimal λ that works well for all parameters in the given range on average.



(a) Improvements of MSE using $\lambda = 1e-6$.

0.17	-0.13	-0.17
0.66	0.43	0.34
0.90	0.75	0.64

10	-0.26	0.03	-0.14	-0.02	-0.09	-0.24	-0.57	-0.29	0.12
20	0.09	-0.23	-0.29	-0.25	-0.28	-0.09	0.04	-0.64	-0.30
30	-0.82	-0.17	-0.40	-0.82	-0.28	-0.02	-0.07	-0.29	-0.51
40	0.17	-0.13	-0.17	-0.61	-0.81	-1.62	-0.22	-0.63	-0.17
50	0.66	0.43	0.34	-0.16	-0.22	-0.34	-0.93	-0.28	-0.44
60	0.90	0.75	0.64	0.53	0.33	0.03	-0.15	-0.11	-0.28
70	0.97	0.92	0.82	0.70	0.61	0.45	0.34	0.37	0.08
	80	90	100	110	120	130	140	150	160

(b) An example of prior range of parameters

Figure: Heat map of effects of regularization using $\lambda = 1e-6$, where $c_1 = 0.5, c_2 = 0.5$, Each entry in the table denotes improvement of MSE over CRLB, i.e., $\frac{CRLB-MSE}{CRLB}$.

Proposed Strategy of Selecting λ's

- For the parameters in a reasonable range (orange region), we compute the average value of improvements, which corresponds to a single value of λ.
- Repeat the above step for all λ's in a reasonable range of λ.
- For all heat maps (corresponding to all λ's), select the one that gives the largest improvement on average.

Conclusions

In brief, we have successfully introduced regularization into NLLS parameter estimation to decrease MSE below both the CRLB and the conventional NLLS analysis.

Future Work

- Given T_{21}, T_{22}, estimate for c_1, c_2.
- All four parameters unknown.
- Other biomedical MRI models.

References I

MJ Box, *Bias in nonlinear estimation*, Journal of the Royal Statistical Society: Series B (Methodological) **33** (1971), no. 2, 171–190.

Yonina C Eldar, *Uniformly improving the cramér-rao bound and maximum-likelihood estimation*, IEEE Transactions on Signal Processing **54** (2006), no. 8, 2943–2956.

Ata Kabán, *Non-parametric detection of meaningless distances in high dimensional data*, Statistics and Computing **22** (2012), no. 2, 375–385.

Dianne P O'leary and Bert W Rust, *Variable projection for nonlinear least squares problems*, Computational Optimization and Applications **54** (2013), no. 3, 579–593.