AMSC 663/664 Proposal Presentation

Analysis of trade-off of variance and bias of parameter estimation in medical imaging

Zezheng Song

Advisor: Dr.Richard Spencer

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

[2]National Institute of Health, National Institute on Aging

October 6, 2020

Content

- Background on parameter estimation in MRI
- Mathematical Formulation
- Numerical Methods
- Preliminary Results
- Validation Methods
- Timelines and Project Schedules

Background

Processes of creation of an magnetic resonance image(MRI)

- Signal generation based on properties of magnetic resonance
- Relaxation processes
- Signal detection
- Encoding spatial information
- Reconstruction of an image from signal

Remark

The project focuses on the second and the third step of the process.

Background

- In magnetic resonance imaging(MRI), an important task is to recover the parameters in the underlying MRI models, which could provide important clinical information. For example, in 1985, Stark found that MR images based on a $\tau 2$-weighted contrast were more effective than CT or radionuclide imaging, in classifying benign or malignant lesions.
- The most common model is the bi-exponential model, i.e.,

$$S(TE; c1, c2, \tau_1, \tau_2) = c_1 \exp\left(-TE/\tau_1\right) + c_2 \exp\left(-TE/\tau_2\right)$$

Motivation of the project

- The bi-exponential model has two terms, which correspond to two different type of molecules. If we can determine those parameters from the signal, then we know how much of each type of molecule is present, which is of great diagnosis value.
- We are developing numerical methods to estimate parameters more accurately.

Definition

In the field of statistics, Cramér–Rao bound gives a theoretical lower bound on the variance of the estimator of deterministic parameter of the model.

- 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 then states that the covariance matrix of $\boldsymbol{T}(X)$ satisfies

$$\mathrm{cov}_{\boldsymbol{\theta}}(\boldsymbol{T}(X)) \geq \frac{\partial \boldsymbol{\psi}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} [I(\boldsymbol{\theta})]^{-1} \left(\frac{\partial \boldsymbol{\psi}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T \tag{0.1}$$

Mathematical Formulation

Limitations of CRB

- CRLB allows evaluation and improvement of variance (by trying to achieve CRLB) of a measurement process.
- Previous work has been focused on providing the Cramér–Rao bound, which gives the lower bound on the variance of the estimators of parameters in the MRI models
- However, in general, we would prefer not to minimize variance, since small variance with huge bias is not desirable.
- Therefore, our goal is to achieve a good trade-off between bias and variance.

Mathematical Formulation

How do we improve it?

- By Yonina Eldar's work[Eld06], the performance of estimators could be improved, provided that we sacrifice some bias
- Box's work [Box71]included important early approximation analysis of bias in nonlinear estimation, with effects of regularization via a Bayesian prior information perspective
- Therefore, we could use Mean Squared Error(MSE) to measure the performance of estimators, by the formula that $MSE(\hat{\theta}) = \mathrm{Var}(\hat{\theta}) + \mathrm{Bias}(\hat{\theta}, \theta)^2$

Mathematical Formulation

Regularized Least Square Problem

$$\hat{\mathbf{p}_\lambda} = \underset{\mathbf{p}}{\operatorname{argmin}} \left\{ \|\mathbf{G}(\mathbf{p}) - \mathbf{d}\|_2^2 + \lambda^2 \left\|\mathbf{L}\left(\mathbf{p} - \mathbf{p_0}\right)\right\|_2^2 \right\} \tag{0.2}$$

- $\mathbf{d}$: noisy data measured in experiment
- $\mathbf{p}$: parameters of the MRI model
- $\mathbf{p}_0$: Priori estimated parameter
- $\mathbf{L}$: weighting matrix L
- $\mathbf{G}$: underlying MRI model, nonlinear in $\mathbf{p}$
- λ: regularization parameter

Advantages

- Makes the potential ill-posed problem stable
- Introduces another free parameter to adjust to a better MSE
- balance the completing goals of minimizing the objective function while controlling the norm of the solution

Problems

- Larger regularization parameter λ gives more stability(less variance), but introduces more bias, vice versa.
- Therefore, we need to choose λ carefully to obtain a desired trade-off between bias and variance.

$$\mathbf{p}_\lambda^* = \underset{\mathbf{p}}{\operatorname{argmin}} \left\{ \|\mathbf{G}\mathbf{p} - \mathbf{d}\|_2^2 + \lambda^2 \|\mathbf{p}\|_2^2 \right\} \tag{0.3}$$

$$= \underset{\mathbf{p}}{\operatorname{argmin}} \left\| \widetilde{\mathbf{G}}_\lambda \mathbf{p} - \widetilde{\mathbf{d}} \right\|_2^2 \tag{0.4}$$

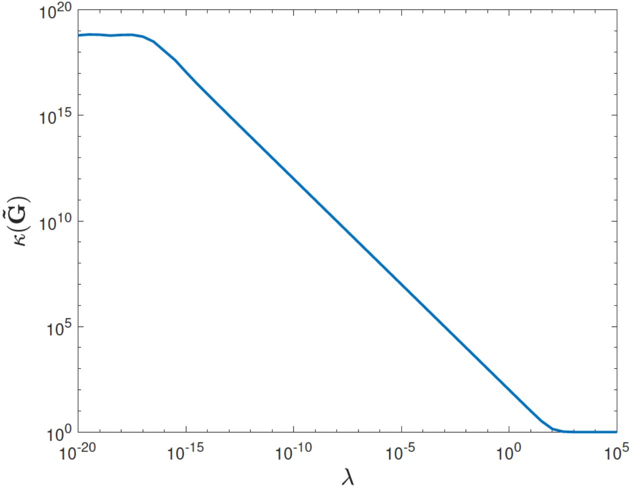
$$= \left(\mathbf{G}^\mathbf{T}\mathbf{G} + \lambda^\mathbf{2}\mathbf{I}\right)^{-\mathbf{1}} \mathbf{G}^\mathbf{T}\mathbf{d} \tag{0.5}$$

where $\widetilde{G}_\lambda = \begin{pmatrix} G \\ \lambda \mathrm{I} \end{pmatrix}$, I is the identity matrix

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

Mathematical Formulation

Relationship of regularization parameter λ and condition number $\kappa(\widetilde{G})$



$$\mathrm{Cov}\left(\mathbf{p}_\lambda^*\right) \approx \left(\left(\widetilde{\mathbf{J}}_\lambda^T \widetilde{\mathbf{J}}_\lambda\right)^{-1} \widetilde{\mathbf{J}}_\lambda^T\right) \mathrm{Cov}(\widetilde{\mathbf{d}}) \left(\left(\widetilde{\mathbf{J}}_\lambda^T \widetilde{\mathbf{J}}_\lambda\right)^{-1} \widetilde{\mathbf{J}}_\lambda^T\right)^T \tag{0.6}$$

$$\approx \sigma_\varepsilon^2 \left(\mathbf{J}^T\mathbf{J} + \lambda^2 \mathbf{L}^T\mathbf{L}\right)^{-1} \mathbf{J}^T\mathbf{J} \left(\left(\mathbf{J}^T\mathbf{J} + \lambda^2 \mathbf{L}^T\mathbf{L}\right)^{-1}\right)^T \tag{0.7}$$

where $\widetilde{\mathbf{J}}_\lambda = \begin{pmatrix} \mathbf{J}\left(\mathbf{p}^*\right) \\ \lambda \mathbf{L} \end{pmatrix}$, σ_ϵ is the standard deviation of data,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

- bi-exponential model

$$S(TE; c1, c2, \tau_1, \tau_2) = c_1 \exp(-TE/\tau_1) + c_2 \exp(-TE/\tau_2)$$

- Stretched exponential:

$$S(b; \alpha, D) = S_0 \exp\left(-(bD)^{\alpha}\right)$$

  , where α is the stretching constant. Arises from some models of restricted diffusion.

- Kurtosis models:

$$S(b; D, K) = S_0 \exp\left(-bD + b^2 D^2 K/6\right)$$

  which is a first-order approximation to non-Gaussian diffusion.

- Gaussian noise, a universal noise model
- Rician noise, common for signal magnitude of MRI images[GP95]

$$p_M(M) = \frac{M}{\sigma^2} \exp\left(\frac{-\left(M^2 + A^2\right)}{2\sigma^2}\right) I_0\left(\frac{A * M}{\sigma^2}\right)$$

- Non-central χ distribution, appropriate for multi-coil acquisition, such as in parallel imaging[BS18]

$$P_\chi(M, A, \sigma, m) = \frac{A^{1-m}}{\sigma^2} M^m \exp\left(-\frac{M^2 + A^2}{2\sigma^2}\right) I_{m-1}\left(\frac{MA}{\sigma^2}\right)$$

where A is the magnitude of underlying noise-free signal; M is the magnitude of the observed signal; m is the number of coils; σ^2 is the noise variance; I_m is the modified mth order Bessel function of the first kind.

Algorithm 1: Monte Carlo Simulation

lambda_vec = logspace($\lambda_{min}, \lambda_{max},$ **number of** λ**s**);
nl = length(**lambda_vec**);
for *n* $\in 1 : nl$ **do**
  lambda = **lambda_vec***(i);*
  sol_vec = **zeros(nrun,4)***;*
  for *j* $\in 1 : nrun$ **do**
    Use optimization solvers to (0.2), and store in **sol_vec***(j,4)*
  end
  Calculate bias, variance, and MSE
end

Mathematical Formulation

Optimization Methods

- Grid Search
- Gradient Descent
- VARPRO[OR13]
- Mesh Adaptive Direct Search
- Levenberg-Marquardt Method

Project Goals

- Study how much amount of regularization we should add, to achieve a desired trade-off between variance and bias? and study if there is a range of λ that makes MSE less than CRLB. If such ranges exist, how confident are we to approximate the true range by our simulation results?
- If such λ exists, study how robust it is with respect to the change within a range of possible parameter values.

Preliminary Work

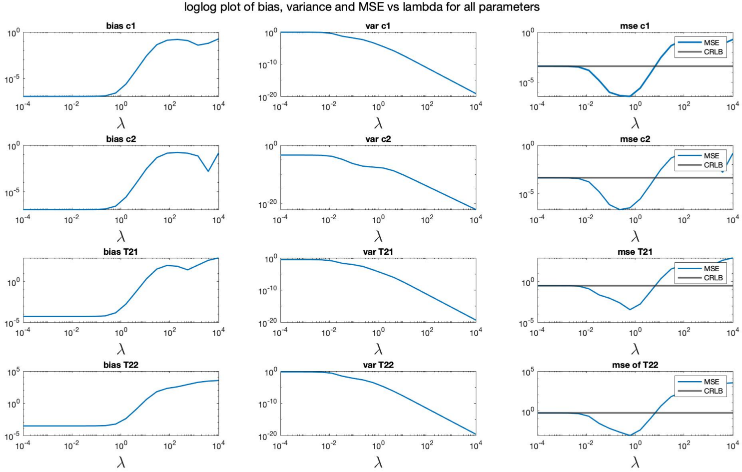
- Studied biexponential model + AWGN,

$$S(TE; c1, c2, \tau_1, \tau_2) = c_1 \exp(-TE/\tau_1) + c_2 \exp(-TE/\tau_2) + \epsilon$$

- For a chosen set of true parameters of the model $(c1, c2, \tau_1, \tau_2) = (0.5, 0.5, 30, 60)$, we implemented Algorithm 1 with the grid search, obtained the plots of bias, variance and MSE vs. λ for each parameter, and compared it with the theoretical results.

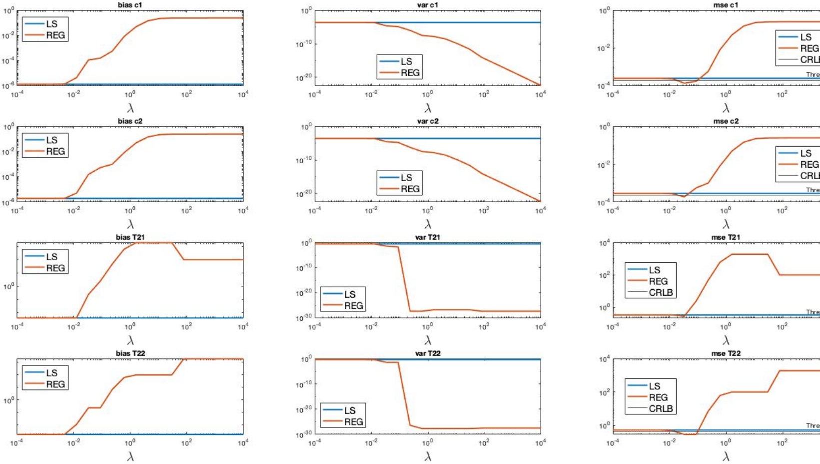
Preliminary Work

Plots of Theoretical Results



Preliminary Work

Plots of Numerical Simulation



Preliminary Work

Table: Range of λ that make MSE below CRLB, calculated theoretically

c1 values	Desired λ range
0.2	(0.0034, 2.8840)
0.3	(0.0020, 4.7121)
0.4	(0.0009, 5.0950)
0.5	(0.0002, 5.0643)
0.6	(0.0004, 4.8688)
0.7	(0.0010, 4.6628)
0.8	(0.0015, 4.4826)

Table: Range of λ that make MSE below CRLB, by grid search

c1 values	Desired λ range
0.5	(0.0230,0.0966)
0.6	(0.0095,0.0142)

Validation

Theory based method

By Box's paper and Spencer's tutorial paper, we are able to calculate the bias and variance, thus MSE. Therefore, it could provide us a validation to our simulation result.

Numerical method

Consider the limiting case(low noise or noiseless), then the performance of the estimator should approach that of the unregularized estimator.

Implementation

- MATLAB 2020A, C++
- High Performance Computing Cluster at UMD

Deliverables

The code will be modularized nicely for the ease of extension, portability and unit testing, with proper documentation

- All MRI models and noise distributions will be written in different MATLAB functions. Data generated by models and noise will be saved in .m files.
- CRLB calculation will be written as MATLAB functions, suitable for different models and noises.
- Solver of optimization problems will be written and tested separately.
- The desired range of regularization parameter will be computed and tested, for different models and noises.

Project Schedules

- Phase 1(October - December): Focus on bi-exponential model and Gaussian noise, exploring different nonlinear non-convex solvers, study the stability of regularization parameter, which will make MSE lower than the corresponding CRLB.
- Phase 2(January and February): With the structure of the codes written, re-do everything with kurtosis, stretched exponential, kurtosis + monoexponential, kurtosis + stretched exponential, all with AWGN.
- Phase 3(March):For bi-exponential model only, extend to Rician noise. Compute the CRLB, analyze the stability of desired ranges of regularization parameter λ.
- Phase 4 (April): If that works, implement Rician noise in other models.

References I

- MJ Box, *Bias in nonlinear estimation*, Journal of the Royal Statistical Society: Series B (Methodological) **33** (1971), no. 2, 171–190.
- Mustapha Bouhrara and Richard G Spencer, *Fisher information and cramér-rao lower bound for experimental design in parallel imaging*, Magnetic resonance in medicine **79** (2018), no. 6, 3249–3255.
- Yonina C Eldar, *Uniformly improving the cramér-rao bound and maximum-likelihood estimation*, IEEE Transactions on Signal Processing **54** (2006), no. 8, 2943–2956.
- Hákon Gudbjartsson and Samuel Patz, *The rician distribution of noisy mri data*, Magnetic resonance in medicine **34** (1995), no. 6, 910–914.
- 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.

Thank you!