An investigation on the relationship between the condition number of design matrices and practical identifiability of parameters

Louis Nass*

Department of Mathematics

Tulane University

New Orleans, LA 70118

December 18, 2019

Abstract

Practical identifiability is established by the confidence interval of a parameter. Here we want to minimize the sum of square residuals as a function of our parameter. With a small parameter confidence intervals, one is able to ensure proper minimization of the total residual. In turn, one can extract the unique parameter that will enhance the model. Here, we will investigate the effect of the condition number of the design matrix on practical identifiability of our parameters. The goal is to establish a relationship between the condition number and practical identifiability in order to further enhance model selection. It is worth considering and investigating how ill-conditioned a matrix can be while retaining practical identifiability. We will perform leave one out profile analysis for various parameters in linear models. For linear models we will consider well-conditioned, ill-conditioned design matrices for square, over-determined, and under-determined systems.

Key Words: Linear models, Condition number, Design matrix, Practical identifiability, Square, Over-determined systems, Under-determined systems, Well-conditioned, Ill-conditioned

*E-mail address: lnass@tulane.edu

Contents

1 Introduction

Given data, $y_i = y(x_i)$, we aim to fit a model in order map the behavior of the data. We do this using both linear and nonlinear models. Fitting these models accurately allows one to use the model as a tool for making predictions attributed to the data. We use models all the time in weather prediction, sports analysis, epidemic mapping etc. Models, both linear and nonlinear, are partially defined by the parameters that are established. Sometimes we use statistical methods to learn which parameters work best, sometimes we use them as tools to enhance predictions. At any rate, we want our parameters to minimize the residuals and be unique. Unique parameters suggest that we have the best model for the data. Problems arise, however, when we are limited in the amount of data we are able to collect. The amount of data collected directly affects our ability to establish unique parameters.

Our goal is to establish a relationship between parameter practical identifiability and the condition number of the design matrix. Having an ill-conditioned matrix will increase the uncertainty, causing problems in attaining our quantities of interest. Hence, we aim to see the effect of the condition number on practical identifiability of parameters.

We first will discuss linear models, particularly regarding the design matrix. We will elaborate on the construction of the design matrix and different types of systems that will require solving. We will discuss the techniques of solving these systems for our parameters, and which matrices will directly effect the error vector.

We will then define parameter identifiability. We will define structural and practical identifiability. We will illustrate the technique used in establishing practical identifiability that will be used in this study.

Next, we will define and discuss the condition number of a matrix. We will distinguish the difference between well-conditioned and ill-conditioned.

Then, we will discuss the experiment. The idea is to create leave one out profile analysis plots for square, over-determined, and under-determined systems that are both well-conditioned and ill-conditioned. We will then make commentary on the results.

The eventual goal of this study is to consider how ill-conditioned a matrix can be while maintaining practical identifiability. This will imply how much data is actually needed to create an accurate model representation of our data.

2 Models

The purpose of a model is to fit data to some function that will allow us to make predictions for data points that we were unable to collect. We define data to be $y_{data}(x_i) = y_{true}(x_i) + \epsilon_i$, where our data at x_i is equal to a true value at x_i plus some error ϵ_i. The behavior of ϵ_i is crucial to our model selection. In this study we will consider $\epsilon \sim N(0, \sigma^2)$ [4], where our error will be identically, independent, and normally distributed with mean 0 and some $\sigma^2 > 0$.

2.1 Linear models

Linear models are defined as $\hat{y} = \mathbf{X}\beta$ where

$$\mathbf{X} = \sum_{j=0}^{p} \phi_j(x) \tag{1}$$

The input vector x will be a matrix as well, but will always have as many rows as observations. We call ϕ_j a basis function. The vector β has dimensions $p \times 1$ and are the coefficients of our basis vectors. The dimensions of $\mathbf{X}$ are crucial for finding the solution to our parameter β.

The first thing that we will do is standardize our linear system. By standardizing our system, we create dimensionless parameters. With dimensionless parameters, we can use our solution techniques for any linear model system, for any type of data.

Here we label $\hat{y} = \mathbf{X}\beta$ and we aim to solve for the $\hat{\beta}$ such that, for $R = y_{data} - \hat{y}$

$$\hat{\beta} = argmin_\beta ||y_{data} - \hat{y}||_2^2 = argmin_\beta R'R \tag{2}$$

We call $\hat{\beta}$, the linear least squares solution. We observe an in-depth proof of this in [3].

2.1.1 Standardization

We standardize our matrix by ensuring that the y_{data} has norm one, and our subsequent matrix has norm one. First we define $||\cdot||$ to be the $L-2$ norm. Then, let S be a $p \times p$ matrix where the j^th column of S be $S_j = \frac{||X_j||}{||y_{data}||}$. Then

$$\mathbf{X}\beta = y_{data}$$
$$\frac{1}{||y_{data}||}\mathbf{X}S^{-1}S\beta = \frac{y_{data}}{||y_{data}||}$$
$$\tilde{\mathbf{X}}\tilde{\beta} = \tilde{y}$$

where $||\tilde{X}||, ||\tilde{y}|| = 1$. From this point forward, we will assume that all matrices and data vectors have been standardized.

There are three different sizes of $\mathbf{X}$. They are square, over-determined, and under-determined.

2.1.2 Square system

Here the dimensions of $\mathbf{X}$ are $p \times p$. If the $Rank(\mathbf{X}) = p$, then there is exactly one unique solution for β. Therefore, we simply solve the system

$$\mathbf{X}\beta = y_{data} \tag{3}$$

If $\mathbf{X}$ is square but the $Rank(\mathbf{X}) = r < p$, we will use the singular value decomposition (SVD) of $\mathbf{X}$ to solve for β. The SVD is as follows

$$\mathbf{X} = UDV' \tag{4}$$

where U and V are orthonormal matrices and $D = diag(\sigma_1, \sigma_2, \ldots \sigma_r, 0, \ldots, 0)$. We will define $D^{-1} = diag(\frac{1}{\sigma_1}, \frac{1}{\sigma_2}, \ldots, \frac{1}{\sigma_r}, 0, \ldots, 0)$ where σ_j for $j = 1, 2 \ldots r$ are the singular values of $\mathbf{X}$. Therefore, we observe that

$$\mathbf{X}\beta = y_{data} \tag{5}$$

$$UDV'\beta = y_{data} \tag{6}$$

$$DV'\beta = U'y_{data} \tag{7}$$

$$V'\beta = D^{-1}U'y_{data} \tag{8}$$

$$\beta = VD^{-1}U'y_{data} \tag{9}$$

However, we do lose information in our matrix due to the rows of zeros in the construction of our D^{-1}.

We will continue to use the SVD to solve for β for our singular systems.

2.1.3 Over-determined system

Here our matrix $\mathbf{X}$ has dimensions $m \times p$ where $m > p$. We call this over-determined because there are more observations than parameters. Because $\mathbf{X}$ has more rows than columns, there are infinite solutions to this linear system. Therefore we will want to find the linear least square solution of the system.

If we have $Rank(\mathbf{X}) = p$, then we use the Moore-Penrose-Psuedo-Inverse (MPP) [2] to find the linear

least square solution for β. Therefore, the system becomes

$$\mathbf{X}'\mathbf{X}\beta = \mathbf{X}'y_{data} \tag{10}$$

and the MPP is $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

For $Rank(\mathbf{X}) = r < m$, we will perform use the SVD to solve the system as used in 5. However, we will take the SVD of $\mathbf{X}'\mathbf{X}$ where we have $\mathbf{X}'y_{data}$ on the right hand side of the system.

2.1.4 Under-determined systems

Here our matrix has dimensions where $m < p$. This occurs when we have more parameters than observations. Inherently, this system also does not have unique solutions, so we will do some manipulation in order to find the linear least square solution. First, we will let $\beta = \mathbf{X}'\eta$, then when we solve the linear least squares of η, this will also minimize β.

For $Rank(\mathbf{X}) = m$, then we can solve

$$\mathbf{X}\mathbf{X}'\eta = y_{data}$$

then $\beta = \boldsymbol{X}'(\mathbf{X}\mathbf{X}')^{-1}y_{data}$ [3].

For $Rank(\mathbf{X}) = r < m$, then again, we use SVD, but for $\mathbf{X}\mathbf{X}'$ to solve for η. Then again $\beta = \mathbf{X}'\eta$. Now that we have solved for our parameters, we will evaluate their identifiability.

3 Identifiability

Even when parameter values minimize the sum of the squared residuals, these solutions are not always unique. When a parameter is not unique, it is not always accurate or useful. When a unique solution for a parameter can be found through parameter estimation, that parameter is said to be identifiable. When multiple values for a parameter can minimize the difference between the model output and the data, that parameter is said to be unidentifiable.

3.1 Structural Identifiability

There are two types of identifiability, structural identifiability and practical identifiability. Structural identifiability occurs when the identifiability comes from the structure of the model. In this case, the identifiable parameters of the model are such that one parameter compensates for one or more other parameters. This causes a manifold of the parameter space along which the parameters can be varied but $R'R$ remains constant

[6]. Here the identifiability is dependent on the model selected. We want to ensure that the parameters we focus on can actually be solved.

Here is an example of structurally unidentifiable parameters. Suppose we use a quadratic polynomial for our model. Observe that

$$\begin{aligned}\hat{y} &= \beta_0 + \beta_1 x + (\beta_2 + \beta_3)x^2 \\ &= \beta_0 + \beta_1 x + \tilde{\beta}_2 x^2\end{aligned}$$

We observe that β_0, β_1 are always structurally identifiable. However, β_2 and β_3 are not structurally identifiable individually. Yet, their sum $\beta_2 + \beta_3 = \tilde{\beta}_2$ is identifiable. This is because there are infinite combinations of β_2 and β_3 that will result in one value for $\tilde{\beta}_2$.

3.2 Practical Identifiability

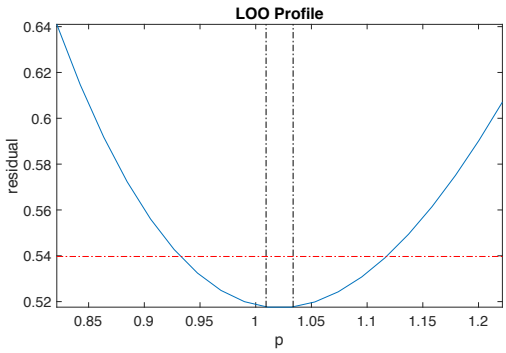
Practical identifiability is a qualitative feature of the parameter that is observed using plotting. Usually, this type of practical unidentifiability comes from either structural unidentifiability or uncertainty in the data that prevents the model from being able to fit the data well [6]. This type of identifiability is intuitive. Is our parameter practical to the model? If we observe a clear minimum for the sum of square residuals, $(R'R)$, within some confidence interval of a parameter, then we say it is practically identifiable. Parameters can have better practical identifiability, this depends on the size of the confidence interval. The smaller the interval, the more practical the parameter is.

There are a number of methods to establish identifiability or unidentifiability of parameters. For the sake of this study, we will only focus on Leave-one-out profile analysis. This is the main test we will use to establish practical identifiability or unidentifiability [7].

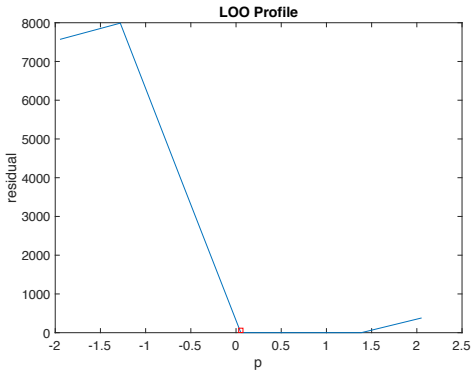
3.2.1 Leave-One-Out profile analysis

Leave-one-out profile analysis (LOO) checks parameters' identifiability across a range of parameter values by comparing the change in $R'R$ with respect to fixing different parameters. This allows assessment of practical identifiability and can be easily implemented.

A benefit of LOO is that it allows assessment of parameters across an extended domain. The process of fixing parameters allows viewing identifiability properties throughout a user defined range of parameter values. A parameter β is then practically identifiable if $R'R$ has a unique minimum within the range of β



(a) The plot above shows a profile for a practically identifiable parameter p. The unique minimum of the blue curve at $p = 1.02$ shows that p is practically identifiable. The two dashed vertical lines represent 95% confidence interval for the parameters found through bootstrapping, and the horizontal dashed line gives the standard error for the residuals.



(b) The plot above shows a profile for a practically unidentifiable parameter p. The existence of a range of minimum values of the blue curve from approximately .05 to 1.4 shows the parameter is practically unidentifiable.

Figure 1: Leave One Out Profile Analysis Examples

values tested (Fig 1a). If there is a range of values for β that all give a minimum residual then the parameter is practically unidentifiable (Fig 1b).

A shortcoming of LOO is that it checks only one parameter at a time. This can create problems because a parameter's identifiability and optimal value is often dependent on the other parameter values. Therefore, even if there is a parameter space where all parameters are identifiable, LOO analysis could be unable to find this point.

4 Condition number

The condition number of a function measures how much the output changes as the input changes. Considering a matrix $\mathbf{X}$ as a system of linear equations, the condition number will give us an indication towards how much the solution $\hat{\beta}$ will change with respect to changes in our result y.

We compute the condition number as follows. Let $\mathbf{X}$ an $m \times p$ matrix. We observe the spectral value decomposition of $\mathbf{X}$ to be

$$\mathbf{X} = UDV'$$

as observed previously in the paper in (5). Again we have $UU' = U'U = I_m$, $VV' = V'V = I_p$, and $D = diag\{\sigma_1, \sigma_2, \ldots, \sigma_n\}$, $n = \min\{m, p\}$, are the singular values of $\mathbf{X}$. The condition number is then

defined as

$$\kappa(\mathbf{X}) = \frac{\sigma_{max}}{\sigma_{min}}$$

where $\sigma_{max} = \max\{\sigma_1, \sigma_2, \ldots, \sigma_n\}$ and $\sigma_{min} = \min\{\sigma_1, \sigma_2, \ldots, \sigma_n\}$ [1].

We refer to the condition number of a matrix $\boldsymbol{X}$ as the norm of the matrix. The notation is as follows

$$||\boldsymbol{X}|| = \kappa(\mathbf{X}) \tag{11}$$

[5].

The condition number is dependent on the singularity of the matrix $\mathbf{X}$. If $\mathbf{X}$ is non-singular, that means that all of the singular values $\sigma_i \neq 0$ for $i = 1, 2, \ldots n$. This will result in a condition number that is not very large. A matrix is considered to be well-conditioned when $\kappa(\mathbf{X}) \approx 1$. On the other hand, if a matrix is singular (or close to singular) then at least one singular value $\sigma_j = 0$ (or $\sigma_j \approx 0$). In this case, the condition number will be undefined or extremely large. This means that the error bound for our solution $\hat{\beta}$ is extremely large, or infinite. Hence, we say that $\mathbf{X}$ is ill-conditioned.

4.1 Effect of condition number on data

Observe for equation (10), if we expand y_{data} in this equation, we have

$$\boldsymbol{X}'\boldsymbol{X}\beta = \boldsymbol{X}'y_{data} \tag{12}$$

$$= \boldsymbol{X}'(y_{true} + \epsilon) \tag{13}$$

$$= \boldsymbol{X}'y_{true} + \boldsymbol{X}'\epsilon \tag{14}$$

$$\tag{15}$$

Here we see that the design matrix ($\boldsymbol{X}$) is directly affecting the error of the system. This in turn will affect the identifiability of the parameter, β, since the condition number $\kappa(\boldsymbol{X})$ will more than likely increase the uncertainty of ϵ.

5 Experiment

In our experiment, we plan on investigating the practical identifiability of a certain parameter in square, over-determined, and under-determined systems.

5.1 Experiment design

First, we will generate some data, using a true function $y_{true} = \boldsymbol{X}_{true}\beta_{true}$ and some error vector $\epsilon \sim N(0, \sigma^2)$ where each ϵ_j will be identical and independent. We will generate some random, standardized matrix $\boldsymbol{X}_{true}$ for the data, that will be size $m \times p$ in dimension, and fit for a fixed β_{true}. We need to ensure in the design of $\boldsymbol{X}_{true}$ that $Rank(\boldsymbol{X}_{true})$ is maximized.

Next, we will consider the the sizes of $\boldsymbol{X}_{true}$. We want to test on square, over-determined, and under-determined systems. Hence we will generate three sets of data to observe the differences of the systems.

Under each system, we will use $\boldsymbol{X}_{true}$ as the well-conditioned design matrix, then we will pick a column $X_{true,j}$ and replace $X_{true,k}$ in $\boldsymbol{X}_{true}$ where $j \neq k$ to create $\boldsymbol{X}_{ill}$.

We will then solve $\hat{\beta}$ for both $\boldsymbol{X}_{true} = \boldsymbol{X}_{well}$ and $\boldsymbol{X}_{ill}$. We will use the MPP techniques for $\boldsymbol{X}_{well}$ and SVD for $\boldsymbol{X}_{ill}$ as described in 2.1. We will then compare the solutions of both parameter estimations to each other and to β_{true}.

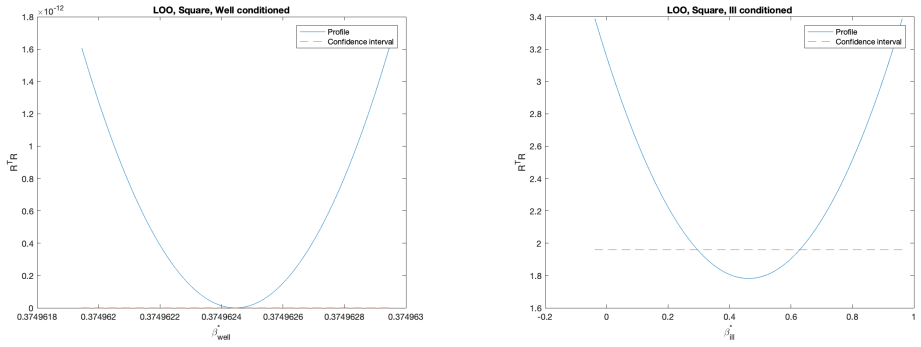
Finally, we will select $\hat{\beta}_i$ from both our well-conditioned and ill-conditioned solved parameters and run LOO profile analysis. We will set the confidence interval of our parameter to be the intersection of the LOO profile and $tol = R'R + (.05)R'R$. This will represent out 95% confidence interval of the parameter. What we aim to see is for smaller confidence intervals for $\hat{\beta}_{well}$ than for $\hat{\beta}_{ill}$.

5.2 Experiment expectations

I expect, in every case, that the well-conditioned systems will have much smaller confidence intervals than the ill-conditioned. I believe this because $R'R$ will be much smaller for the well-conditioned systems than the ill-conditioned systems. With a smaller $R'R$ obviously the tol, defined in section (5.1), will be smaller as well. Additionally the $\hat{\beta}_{well}$ will inherently have less error than $\hat{\beta}_{ill}$.

6 Results

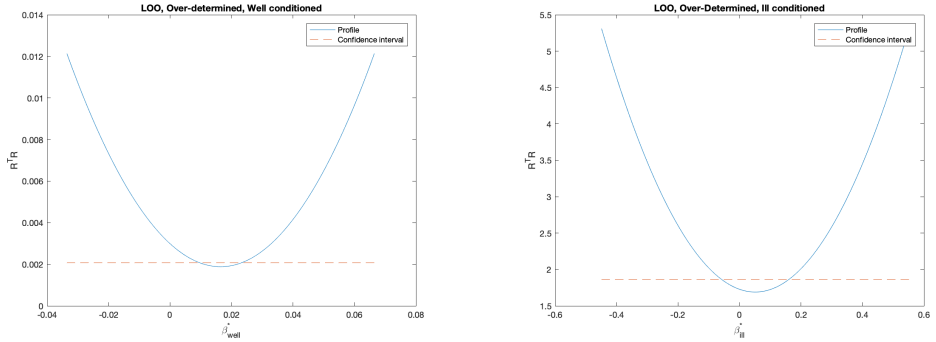
Here we will observe the LOO profile analysis plots generated from the experiment. We will compare the well-conditioned and ill-conditioned systems of the square, over-determined, and under-determined systems respectively.



(a) LOO profile for well-conditioned square matrix. Here we see that there is exactly one solution that minimizes $R'R$.

(b) LOO profile for ill-conditioned square matrix. Here we see that the confidence interval for this system is about $\beta^*_{ill} \in [0.37, 0.61]$

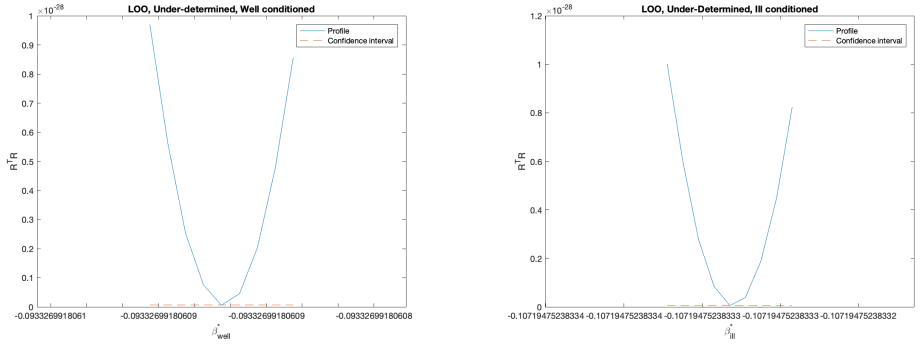
Figure 2: LOO profile analysis figures for square system



(a) LOO profile for well-conditioned square matrix. Here we see that the confidence interval is about $\beta^*_{well} \in [0.019, 0.025]$.

(b) LOO profile for ill-conditioned square matrix. Here we see that the confidence interval for this system is about $\beta^*_{ill} \in [-0.05, 0.17]$

Figure 3: LOO profile analysis figures for over-determined system



(a) LOO profile for well-conditioned square matrix. Here we see that the confidence interval is extraordinarily small and hard to see.

(b) LOO profile for ill-conditioned square matrix. Here we see that the confidence interval also too small to observe.

Figure 4: LOO profile analysis figures for under-determined system

6.1 Square system

In the square system, we generate a 5×5 design matrix $\boldsymbol{X}$, as described in 5.1. For the square system, we will observe the $\kappa(\boldsymbol{X})$. For each system they were

Square System	
Well-Conditioned System $\kappa(\boldsymbol{X})$	Ill-Conditioned System $\kappa(\boldsymbol{X})$
2.69E+01	2.49E+18

From viewing (Fig 2a) and (Fig 2b), we are further asserted from our initial assumptions. A full rank, square matrix will have a unique $\hat{\beta}$ that will solve the system. Even with some error ϵ, we see that the confidence interval is a single value or extremely small.

For an ill-conditioned matrix, we see that there is a clearly a larger confidence interval, however we are somewhat happy because the parameter in this system did retain structural identifiability. The interval is approximately $\hat{\beta}_{ill} \in [0.37, 0.61]$, and that is a relatively small interval.

6.2 Over-Determined system

In the over-determined system, we generate a 10×5 design matrix $\boldsymbol{X}$, as described in 5.1. For the over-determined system, we will observe the $\kappa(\boldsymbol{X}'\boldsymbol{X})$. For each system they were

Over-Determined System	
Well-Conditioned System $\kappa(\boldsymbol{X}'\boldsymbol{X})$	Ill-Conditioned System $\kappa(\boldsymbol{X}'\boldsymbol{X})$
8.90E+00	2.38E+32

Here we see that the condition numbers are both much larger than the square system. Additionally, from (Fig 3a) and (Fig 3b), we are seeing clear confidence intervals in both systems

We see for $\hat{\beta}_{well}$ the confidence interval is approximately $[0.019, 0.025]$. This is clearly larger than the square matrix.

For $\hat{\beta}_{ill}$ the interval is $[-0.05, 0.17]$. This again is larger than the interval for the well-conditioned system.

6.3 Under-Determined system

In the under-determined system, we generate a 5×10 design matrix $\boldsymbol{X}$, as described in 5.1. For the under-determined system, we will observe the $\kappa(\boldsymbol{X}\boldsymbol{X}')$. For each system they were

Under-Determined System	
Well-Conditioned System $\kappa(\boldsymbol{X}\boldsymbol{X}')$	Ill-Conditioned System $\kappa(\boldsymbol{X}\boldsymbol{X}')$
1.20E+01	1.81E+01

Interestingly enough, the condition numbers for $\boldsymbol{X}\boldsymbol{X}'$ in both systems are relatively small. They are similar to each other in size. This becomes apparent when we investigate (Fig 4a) and (Fig 4b). We see that in both figures, it appears that both parameters are almost exactly unique, where the confidence intervals are minuscule. However, we do recognize that the ill-conditioned parameter is converging to a different value than our well-conditioned value.

Hence we recognize that the ill-conditioned parameter has a small interval, but does not have the correct value.

7 Conclusion

Considering the figures created. There are clear distinctions between well-conditioned and ill-conditioned systems and well as differences between square, over-determined, and under-determined systems.

7.1 Main take-away

From looking at the figures on page 11, we see that ill-conditioned systems consistently have larger confidence intervals for our parameters. This suggests that having an ill-conditioned matrix does indeed reduce practical identifiability.

What is interesting is to compare the results of square, over-determined, under-determined systems. In both the square and over-determined systems, the ill-conditioned design matrix yielded confidence intervals for our parameter that intersected with the confidence intervals of the well-conditioned system. This suggests that the error caused by the condition number of of design matrix stretched the interval more than it shifted. Hence, in the over-determined system, we retain accuracy but lose precision in finding the true parameter value.

For the under-determined system, we observe that the confidence intervals are almost non-existent. The parameters were mapped effectively to a single value with very little variance. However, the well-conditioned system yielded a completely different value than the ill-conditioned system. This suggests that although we gain precision, we lose accuracy in this system.

7.2 Further Study

To expand on this study, the next step would be to consider the linearization of nonlinear models. The point of this experiment is to test the condition number of some matrix. In order to consider non-linear models, one would need to find a way to transform the model to a linear system. The design matrix is the key to this experiment, and it would not apply to a linear model that can not be linearized.

References

[1] Wikipedia contributors. Condition number — Wikipedia, the free encyclopedia, 2019. Online; accessed 2019-12-10.

[2] Wikipedia contributors. Moore-Penrose inverse — Wikipedia, the free encyclopedia, 2019. Online; accessed 2019-11-16.

[3] Onur Danaci, Harley Hanes, Elliot Hill, Louis Nass, Noah Rahman, Yi Tang, Zheng Wang, Helen Weierbach, and James M. Hyman. Fitting linear and nonlinear models to data and quantifying their uncertainty. *Tulane Math Archive*, pages 1–53, 2019.

[4] Per Christian Hansen, Víctor Pereyra, and Godela Scherer. *Least squares data fitting with applications*. JHU Press, 2013.

[5] Nicholas J. Higham. *Accuracy and stability of numerical algorithms*. Siam, ii edition, 2002.

[6] A. Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, and J. Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25:1923–1929, 2009.

[7] Ralph C. Smith. *Uncertainty Quantification: Theory, Implementation, and Applications*. SIAM, 2014.