École Polytechnique

BACHELOR THESIS IN COMPUTER SCIENCE

Investigating The Loss Landscape For ODE Parameter Inference

Author:

Amine Roudani, École Polytechnique

Advisor:

Gleb Pogudin, LIX

Academic year 2023/2024

Abstract

This thesis explores the landscape of the loss function in parameter estimation problems, in the context of differential equation models with the exponential growth model under examination. Our approach based on transforming the loss function into a polynomial in order to use algebraic techniques such as Gröbner bases and Homotopy Continuation. The study explores how various factors such as noise, number of datapoints, and removal of datapoints, influences the number of distributions of minima in the loss function. Furthermore the study also extends into a two-dimensional model. This work not only enhances the theoretical framework for parameter estimation in differential equation models but also provides a solid foundation for future research in extending these methodologies to higher-dimensional and more complex models.

Contents

1 Introduction **4**
- 1.1 Set-up . . . 4
- 1.2 Gröbner Bases . . . 5
- 1.3 Homotopy Continuation . . . 5

2 Experiments **6**
- 2.1 Methodology . . . 6
- 2.2 How Noise Level Affects Number of Roots . . . 8
- 2.3 How Noise Level Affects Number of Minimas . . . 9
- 2.4 How Number of Datapoints Affects Number of Minimas . . . 10
- 2.5 How Removing Datapoints Affects Number of Minmia . . . 11
- 2.6 Investigating Multi-Minima Cases . . . 12
- 2.7 Investigating the Multi-Dimensional Case . . . 13

3 General Cases **15**

4 Conclusion **16**
- 4.1 Summary of Major Findings . . . 17
- 4.2 Contributions to Computer Science . . . 17
- 4.3 Suggestions for Future Research . . . 17

5 References **19**

A Appendix **20**

1 Introduction

1.1 Set-up

Differential equations are fundamental in modeling the dynamics of natural phenomena, providing insights into various scientific and engineering disciplines. The process of estimating model parameters to fit a dataset is usually approached as follows: for any fixed parameter values, one defines a loss function as the deviation of the solutions with the given parameter values from the observed data [4, 3]. Then this loss function is somehow minimized. However, regardless of the method chosen, an important issue is whether the found minimum will correspond to the ground truth values. The optimization landscape is often marred by the presence of multiple local minima, complicating the identification of true parameter values. In other words, we often do not know for a fact whether the minima found is local or global, and we do not know what distribution the minimas follow.This challenge, pronounced even in models described by the simplest differential equations, will be explored in this paper. The goal of this thesis is to focus on differential equations with known analytical solutions, use the available explicit formulas to write down the explicit expressions for the loss function, and analyze the number and distributions of local minima of the loss. We begin with a parametric model

$$x' = f(x, \alpha), \tag{1}$$

where x is a function of time t, and α represents the scalar parameter of the system. In this report, the focus is on the following differential equation

$$x' = \alpha x, \tag{2}$$

This differential equation can easily be solved analytically:

$$x(t) = Ae^{\alpha t}, \tag{3}$$

where A is the initial condition. sAs mentioned above, when estimating the model parameters A and α, we minimize a loss function. In this report, we use the so-called Sum of Squared Error as the loss function. In general, given a set of observed values x_i and a set of predicted values $\hat{y}_i$ derived form a model over n observations, the Sum of Squared Error is defined as:

$$SSE := \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{4}$$

In this case, with the model defined in equation 2, and given a known data set $(x_i, t_i)_i$, then the Sum of Squared Error can be written as

$$SSE := \sum_{i} (Ae^{\alpha t_i} - x_i)^2 \tag{5}$$

The key innovation in our approach is the transformation of SSE into a polynomial form. In general, this can be done for linear differential equations with known analytical solutions. Of course, we make the important assumption that t_i's are integers so that the SSE transforms indeed into a polynomial. This transformation of the SSE into a multivariate polynomial is done by introducing the following definition

$$\beta = e^{\alpha}. \tag{6}$$

We finally obtain that the SSE can be written as a polynomial in A and β, with the time points t_i acting as the exponents of the polynomial

$$SSE := \sum_{i} (A\beta^{t_i} - x_i)^2 \tag{7}$$

This reformulation of the Sum of Squared Error as a polynomial function enables us to compute the gradient of the SSE as another polynomial function, which facilitates the application of algebraic methods, such as Gröbner bases and Homotopy Continuation, to determine the exact number of roots of the gradient function(and consequently, the exact number of minima of the SSE itself). This allows us to analyze the number and distributions of local minima of the loss function, which is the goal of this thesis.

Furthermore, by extending this analysis to a general parametric differential equation $x' = f(x, a)$ and elaborating on the process of deriving the SSE in its general form, we outline how the search for local minima essentially reduces to solving a system of polynomial equations obtained by taking the gradient of the SSE. Overall, this thesis merges analytical and computational strategies to enhance our comprehension of parameter fitting in differential equation models, contributing both theoretical insights and practical methodologies to the scientific community.

1.2 Gröbner Bases

Gröbner bases were introduced by Bruno Buchberger in 1965, are a foundational tool in computational algebraic geometry [2]. Moreover, in order to define Gröbner bases, we must first define *ideals*

Definition An *ideal* is a special subset of a ring. For a given ring R, an ideal I is a subset of R that has the following properties:

1. It is non-empty.
2. For any two elements $a, b \in I$, the sum $a + b$ is also in I.
3. For any element $a \in I$ and any $r \in R$, the products ra and ar are also in I.

A Gröbner basis for an ideal I in a polynomial ring $R = k[x_1, \dots, x_n]$ is a set of polynomials with specific properties that simplify the process of solving polynomial systems by providing a systematic way for solving systems of polynomials. Formally, a set of polynomials $G = \{g_1, g_2, \dots, g_t\}$ is a Gröbner basis for an ideal I if and only if the leading term of every polynomial in I, with respect to a chosen monomial order, is divisible by the leading term of some polynomial in G. This definition relies on the concept of monomial orders, which are ways to provide a total order on the set of monomials in R.

In the context of this thesis, Gröbner bases are used to solve for the roots of the gradient of the multivariate polynomial in β and A that defines the SSE (7). Specifically, we use the *elimination property* of Gröbner bases. This is a critical feature that facilitates solving systems of polynomial equations by systematically reducing the number of variables in each equation. Given a Gröbner basis for an ideal with respect to a lexicographic order where variables are ordered as $x_1 > x_2 > \dots > x_n$, the elimination property guarantees that for every k from 1 to n, if there is a polynomial in the basis that involves only the variables $x_1, x_2, \dots, x_k$ present in the ideal, then it will exist. We use the elimination property of Gröbner bases to transform the system of equations obtained by setting the gradient to zero. This gives us two polynomials: one single-variable polynomial in β, and another polynomial that is linear in A. The degree of this polynomial in β gives the number of roots directly, and we can then apply Strum's Algorithm to count the number of *positive* roots since by definition $\beta = e^{\alpha} > 0$. Furthermore, Gröbner bases significantly simplifies the process of solving the system: once β is solved from the single-variable polynomial , the solutions can be back-substituted into the other polynomial to find A. This approach is crucial for identifying the roots of the polynomial, corresponding to the local minima of the SSE, thereby aiding in the parameter estimation problem for differential equations.

1.3 Homotopy Continuation

Homotopy continuation is an alternative option to Gröbner bases for solving systems of polynomial equations, particularly useful in our case for finding the roots of the Sum of Squared Errors (SSE) in our parameter estimation problems. This method iteratively tracks the solutions of a system of equations from a starting point, where solutions are known or easy to find, to the target system whose solutions are sought. The essence of homotopy continuation lies in constructing a homotopy between a simple system of equations, whose solutions are readily available, and the complex system of interest. Consider the polynomial reformulation of the SSE as expressed in Equation (7). The goal is to find the roots of this polynomial system, corresponding to the parameter values that minimize the SSE. The process begins with defining a homotopy

$$H(x, \lambda) : \mathbb{C}^n \times [0, 1] \to \mathbb{C}^n \tag{8}$$

which connects a start system $G(x) = 0$ with known solutions to the target system $F(x) = 0$ (the gradient of the SSE polynomial). This is often done as:

$$H(x, \lambda) = (1 - \lambda)G(x) + \lambda F(x), \tag{9}$$

where λ varies from 0 to 1. At $\lambda = 0$, $H(x, 0) = G(x)$ is the starting system, and at $\lambda = 1$, $H(x, 1) = F(x)$ is the target system. Then, to solve $F(x) = 0$, we start with the known solutions of $G(x)$ at $\lambda = 0$ and incrementally track these solutions as λ moves towards 1. This path tracking is typically done using numerical continuation methods, adjusting λ in small steps and solving $H(x, \lambda) = 0$ at each step until reaching the solutions of $F(x)$.

Homotopy continuation is particularly effective in dealing with systems that have numerous variable as well as numerous solutions, including complex solutions, and offers a robust way to navigate through the complex solution landscape of polynomial equations. Its application in this thesis leverages its strength to systematically explore the parameter space defined by the SSE, aiming to find all possible roots that correspond to local minima. This comprehensive approach ensures that the global minimum can be identified among the potential multiple local minima, overcoming the limitations often faced by traditional optimization methods in complex loss landscapes.

By using homotopy continuation, we avoid the computational difficulties associated with the exponential complexity of Gröbner bases which comes into play when more variables are introduced, thus providing a numerically stable method to explore the entirety of the solution space. This method's integration into the analysis of differential equations with known analytical solutions presents a novel pathway to addressing the challenges of parameter estimation, highlighting the fusion of algebraic techniques and numerical methods in advancing the field of mathematical modeling.

2 Experiments

In this section, we first outline our methodology used. Then we follow with a deep dive into a series of experiments and simulations aimed at exploring the minima landscape. The experiments include investigations into the effects of noise levels on root identification, the impact of data quantity on minima stability, and extends the realm of multi-dimensional models. For full overview and comprehensive insight, the entire collection of experiments is accessible in our GitHub Repository. Moreover, key parts of code critical to our discussion are directly linked throughout this section for ease of reference.

2.1 Methodology

The methodology used in this study is designed to process a given dataset as its input and yield the exact count of minima for the Sum of Squared Errors (SSE) polynomial, referenced in equation (5), as its output. The objective is to execute this process with speed, efficiency, and accuracy. The method is detailed through a series of steps that encompass key considerations and decisions tailored to the experimental context. By varying a single variable while holding others constant, the approach aims to identify and analyze clear causal relationships within the data.

1. **Generating Data** The initial phase is data generation, where the input parameters include the desired number of data points, the initial condition A, and the exponent α. These parameters are selected carefully to ensure that data values remain reasonable, avoiding computational slowdowns that come from excessively large exponential data in subsequent experiments. Upon setting these parameters, data points are generated and then subjected to Gaussian noise, with a noise level parameter chosen within the interval $[0, 1]$. The noise is proportional to the data values and simulates real-world data imperfections. The output of this step is a dataset ready for further analysis, incorporating both the underlying trend dictated by A and α, and the added variability from the noise.

2. **Computing Gröbner Bases** This goal of this step is to take the equation $\lambda SSE = 0$ and to transform this system into a more solvable form. The use the computation of Gröbner bases facilitates the conversion of this system into two equivalent polynomials that are simpler to solve. This computation is executed using the Python SymPy library, which requires approximating the real values x_i in our dataset to rational numbers for efficiency and compatibility reasons. Despite this approximation, the noise inherent in our data ensures the reliability of our results. The output from this step is a pair of equivalent polynomials, obtained through the Gröbner basis, which are then ready for the root-finding process to identify potential minima locations. Moreover, it is essential to underline that the t_i values are integers, a condition specified during the introduction of this report.

3. **Solving for** β: The next step is to solve for β in the single-variable polynomial computed using the Gröbner basis. Recognizing that $\beta := e^{\alpha}$ means that β must be positive, so only positive roots are considered. The process unfolds as follows:

   i Before drectly computing the roots, Sturm's theorem is used to determine the count of positive roots present in the polynomial. Sturm's theorem, a fundamental principle in algebra, provides a reliable method to determine the exact number of real roots within a specific interval without actually solving the equation[1]. This step is crucial as it sets a clear target for the root-finding process, while ensuring our focus remains solely on positive roots, in alignment with the definition $\beta := e^{\alpha}$.

   ii Once the precise number of positive roots established, we start the root-finding process with the SymPy library's `solve()` function. This method, chosen for its efficiency and simplicity, if successful in identifying all positive roots, marks the end of our search. If any roots remain undiscovered, we move onto the next step.

   iii The second approach for determining the positive roots is Newton's method. Random initial guesses within a predetermined range are the starting point of this iterative method. The algorithm adapts dynamically for each root we want to find, iterating up to a predetermined maximum till we find a solution. By comparing each root to previously discovered roots with a tolerance level, duplicate discoveries are avoided and each root's uniqueness is guaranteed. If this approach is unable to converge on a root in the allotted number of tries, it indicates that the root is no longer being searched for and moves on to make sure no other possible solution is missed. This method locates the roots while making sure they are distinct and constructive.

   iv Should both of the above methods prove insufficient, we resort to the Grid Search technique, detailed in step 6. This approach is our safeguard against numerical challenges that might obscure potential solutions, ensuring no stone is left unturned in our search for positive β values.

   This structured and systematic approach not only ensures to always find positive β values (if they exist), but also ensures that the methodology adheres to the underlying mathematical principles guiding our analysis.

4. **Computing the Parameters**: After having identified the positive β values, the subsequent phase involves deducing the exponential model parameters. The process now uses on the Gröbner basis's other

polynomial, which is linear in A, as discusses in 1.2. By substituting the β values found in the previous step into this polynomial, we can extract the corresponding initial conditions A. The computation of α values follows directly, simply using the definition $\beta = e^{\alpha}$. This procedure translates the abstract β values into the actual model parameters A and α.

5. **Classifying the Extrema**: The focus of this stage is on determining whether the derived parameter pairs (A, α) constitute minima of our model (as they may also be saddle points or maximas). This involves the symbolic computation of the general Hessian matrix, into which each parameter pair is then substituted to assess the nature of the extrema. Usually, one would compute all the eigenvalues and check that they are positive. However, since we are dealing with 2x2 symmetric matrices, it suffices to check the signs of the determinant and the trace to determine whether given point is a minima. Computing determinant and trace is more efficient and quicker. This method, grounded in classical optimization theory, ensures a rigorous evaluation of each potential extrema. However, despite the thoroughness of this approach, there are cases where no minima are identified due to numerical errors. In such scenarios, the methodology transitions to the Grid Search 6 step as a contingency measure to locate the minima. Otherwise, should minima be confirmed, the process advances to the final iteration step 7, thus progressing towards the comprehensive analysis of the model's extremal points.

6. **Grid Search** : This critical step is designed to address any numerical challenges that have emerged in the model thus far. Conducted within the anticipated minima regions of the A - α parameter space, the grid search methodically and systematically computes the loss for each coordinate point within a predefined grid. This exhaustive search aims to identify the point with the minimal loss value, thereby nominating it as a *potential* minima. Then key distinction is made based on the location of this minimal point within the grid: if located *inside* the grid, it is likely indicative of a true minima within the loss function's landscape. Conversely, a minimal point positioned on the grid's *periphery* (i.e. edges) suggests an asymptotic approach towards zero of the gradient function, making it unlikely to be a true minima.

7. **Iterations** : The final phase of this process is a repetitive loop that encapsulates the whole process and loops over the entire process from data generation to extrema classification. This iterative cycle is designed to refine our results thus enhancing the precision of the model's identification of minima. This is a systematic approach aiming to improve results, ensuring robustness and accuracy in the findings.

This method enables us to change the data parameters one by one, while keeping the others constant, to see the effect of the number and distribution of minimas.

2.2 How Noise Level Affects Number of Roots

This initial exploration examines the effect of data noise on identifying the positive roots in the SSE polynomial equation. We methodically vary the noise level and use Sturm's theorem to count the exact number of positive real roots, while maintaining all other variables constant. This way we can deduce any trends and causal effects. This foundational experiment paves the way for further analysis into how variations in noise level influence the detection of minima, which will be the focus of our subsequent experiment. The results of this experiments are captured in a comprehensive table1 summarizing the mean, standard deviation, and maximum number of positive roots identified across various noise levels. The full experiment is accessible via this Jupyter Notebook.

Table 1: Number of Positive Roots for Various Noise Levels

Noise Level	Mean Roots	Std Dev	Max Roots
0.0	1.0	0.0	1
0.1	1.0	0.0	1
0.5	1.3	0.2	2
1.0	1.2	0.4	2
1.5	1.2	0.4	2
3.0	1.4	0.5	3
5.0	1.6	0.6	3
10.0	1.7	0.6	3
20.0	1.8	0.7	3
30.0	1.8	0.7	3
40.0	1.9	0.7	4
70.0	1.9	0.7	3
100.0	2.0	0.7	4
200.0	1.9	0.7	4
300.0	2.2	0.8	4
400.0	2.1	0.8	4
500.0	2.1	0.8	4

When analyzing the data in the table above, we notice a relationship between the noise level in the dataset and the number of positive roots identified through our computational framework. More specifically, we observe that as the noise level increases, so does the average number of positive roots. This implies a direct correlation between data noise and the complexity of parameter estimation. Furthermore, this trend highlights the challenges posed by noisy data in accurately identifying model parameters. It also highlights the importance of robust computational techniques in managing such uncertainties. Furthermore, we can hypothesize the following conjecture:

Proposition 2.1. *Consider a dataset precisely conforming to an exponential growth model as defined in 2, without any form of noise. In this scenario, the Sum of Squared Errors (SSE) function is guaranteed to exhibit one single* ***positive root****.*

Remark 2.1. *The unique positive root identified under the noiseless condition, as described in Proposition 2.1, does not result from the convexity of the Sum of Squared Errors (SSE) function. Examination of the Hessian matrix for any given dataset adhering to the exponential model reveals that it is not positive (semi-)definite, indicating the minima arises from the model's intrinsic properties rather than convex optimization conditions.*

To conclude our analysis of how the noise level impacts the number of minimas, a natural progression emerges: 'How many of these positive roots correspond to actual minimas?' This is exactly what will be investigated in the next subsection.

2.3 How Noise Level Affects Number of Minimas

This experimental investigation looks at the relationship between noise levels and the emergence of minimas within the Sum of Squared Errors (SSE) framework. We change the level noise introduced into the system in an attempt to uncover patterns, thresholds, and insights that delineate the behavior of minimas in response to increasing disturbances. The full experiment is accessible via this Jupyter Notebook.

Table 2: Summary of Minima Analysis

Noise Level	Mean Minima	Std Dev	Max Minima
0.0	1.0	0.0	1
0.1	1.0	0.0	1
0.5	1.0	0.0	1
1.0	1.0	0.0	1
1.5	1.0	0.0	1
3.0	1.2	0.4	2
5.0	1.1	0.3	2
10.0	1.1	0.3	2
20.0	1.1	0.5	2
30.0	1.3	0.5	2
40.0	1.2	0.6	2
70.0	1.4	0.7	2
100.0	0.9	0.7	2
200.0	1.3	0.5	2
300.0	1.0	0.8	2
400.0	1.0	0.6	2
500.0	1.3	0.5	2

The results summarized in Table 2 show some key trends. Firstly, similar to the previous experiment, we have the presence of at least one minima across all examined noise intensities. Moreover we also have what seems to be a sort of inflection point at around noise level 3.0 as we get at least 2 minimas after that. Again, this trend, where the mean number of minima seems to rise with noise, underscores the important relationship between noise and the SSE landscape's complexity. Such findings are key in understanding the robustness of exponential growth models against data perturbations, aligning with our thesis' inquiry into the resilience and adaptability of these models under real-world experiments and scenarios.

Furthermore, we can now build upon Propositions 2.1 from las section as follows

Proposition 2.2. *Take a dataset that precisely follows an an exponential growth model defined in 2, without any form of noise. In this scenario, the Sum of Squared Errors (SSE) function is guaranteed to exhibit one single **minima***

2.4 How Number of Datapoints Affects Number of Minimas

This section explores how variations in the number of data points impacts the number and distribution of minima. Given the variability of dataset sizes in real-world experiments, from small-scale observations to extensive collections of data, understanding this aspect is very important for accurate modeling and analysis across different contexts. We change that the dataset size, while keeping all else constant, in order to investigate the influence of data-point quantity on the stability and identification of minima. The full breakdown of this experimental exploration can be found in thisJupyter Notebook.

Table 3: Summary of Number of Datapoints Analysis

Number of Datapoints	Mean Minimas	Std Dev	Max Minimas
2	1.0	0.0	1
3	1.0	0.0	1
4	1.0	0.0	1
5	1.0	0.0	1
6	1.0	0.0	1
7	1.0	0.0	1
8	1.0	0.0	1
9	1.0	0.0	1
10	1.0	0.0	1
11	1.0	0.0	1
12	1.0	0.0	1
13	1.0	0.0	1
14	1.0	0.0	1
15	1.0	0.0	1
16	1.0	0.0	1

Looking at the results summarized in table 3, we see a clear patter. The mean number of minima remains constant at one at all numbers of data-points, with absolutely no variability (standard deviation of 0.0). This uniformity and consistency suggests that within the constraints of our model and the experimental setup, the quantity of data points does not influence the emergence or stability of minima. Such findings underscore the robustness of the model in identifying a singular minima across varying dataset sizes, highlighting its potential applicability and reliability in diverse real-world scenarios. Further, we propose the following hypothesis:

Proposition 2.3. *Given a dataset that follows exactly an exponential growth model defined in 2, with little-to-no noise. In this scenario, the Sum of Squared Errors (SSE) polynomial is guaranteed to exhibit one and only one minima.*

Building on this section, a logical progression leads us to question the effects of irregular data distribution, particularly the impact of non-uniform time-stamp intervals on the identification of minima. The next section delves into this exact scenario.

2.5 How Removing Datapoints Affects Number of Minmia

In real-world scenarios, data collection seldom occurs at uniformly distributed time intervals. Due to the nature of the experimental setup, inherent constraints of the measurement tools, and the dynamics of the system under study, we will often deal with varied time-stamping for data points. Non-uniform distribution of time-stamps may suggest that we may have more complexities in the analysis, particularly in the accurate identification of system parameters and the subsequent localization of minima in the parameter space.

When running this experiment, we use the same method as before but with a small modification. Right after the generation of data step in the methodology 2.1, we introduce a new intermediate step wherein data points are randomly deleted before we move to the computation of the Gröbner basis and subsequent analytical processes. This new step allows us to observe how the number and distribution of minimas is affected by datasets compromised by arbitrary removals. The full experimentation conducted, as detailed in this Jupyter Notebook, reveals that despite the deliberate random omission of data points from datasets, the analysis consistently identified an unchanged number of minima. This result is the same as in the previous section, and it leads us to hypothesize the following proposition.

Proposition 2.4. *Given a dataset that follows exactly an exponential growth model defined in 2, with little-to-no noise. In this scenario, deleting data-points from the original noiseless dataset was no effect on the number and distribution of minima.*

2.6 Investigating Multi-Minima Cases

So far, our exploration has primarily focused on identifying the number and distribution of minima within the Sum of Squared Errors (SSE) polynomial for datasets following exponential growth model. However, now we will move beyond just identifying the minimas. We turn our attention to the nuanced landscape of multi-minima cases, specifically examining how the value of the SSE at different minima compares. This comparison is not only of theoretical interest but also bears significant practical significance, as it informs the selection of optimization strategies and highlights potential implications in settling for suboptimal solutions. We begin with a simple example. The graph below depicts a dataset and includes two exponential curves that represent the two minima identified using the previously outlined methodology 2.1. Figure 1 below illustrates a scatter plot of the dataset along with the two exponential curves. Each curve corresponds to a distinct set of parameters from the minimas, showcasing how they fit the data differently.

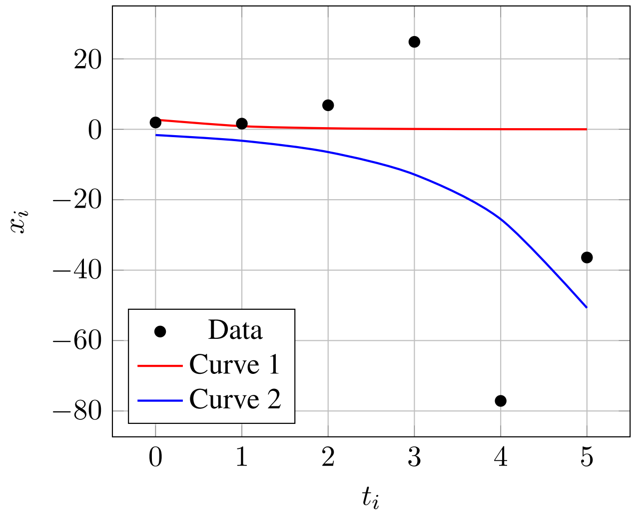

Figure 1: Scatter plot of the dataset with two exponential curves.

The table 4 below lists the parameters of the exponential curves and their corresponding sum of squared errors (SSE).

Initial Value (A)	Alpha (α)	SSE Error
2.7	-1.1	7938
-1.6	0.7	4502

Table 4: Parameters of the exponential curves and their SSE errors.

As can be seen in Table 4, the differences in the SSE values for different minima are important, with one minima demonstrating an SSE nearly double that of the other. This variance prompts a deeper dive into the SSE differences across minima in our model's context.

In an attempt to answer this question, we conduct a new experiment using the methodology outlined in 2.1. This time, we focused only on instances where exactly two minima were identified, and computed the SSE difference between them. This procedure was repeated over 100 times to ensure the reliability of our findings. The comprehensive details of this experiment are accessible in the accompanying Jupyter Notebook. The aggregate results of our investigation are summarized below in Table 5, providing an analysis of SSE variations between dual-minima scenarios.

Mean ΔSSE	Max ΔSSE	Min ΔSSE	Std. Dev.
2.37e04	9.51e04	3.55e-11	3.41e04

Table 5: Differences in SSE

As can be observed in table 5, the differences in the SSE in the cases of two minima can be very large at times. The cases where the SSE delta approaches zero means that both of the identified minima offer comparably optimal solutions. Conversely, when we have significant SSE disparities, this highlights the risk of optimization processes converging on suboptimal minima. The latter outcome underscore the importance for cautious selection of optimization strategies to avoid "locking in" to less desirable solutions due to misleading initial conditions or inadequate analytical methods.

2.7 Investigating the Multi-Dimensional Case

In this subsection, we expand our analysis to a two-dimensional model to examine its impact on the number and distribution of minima. We define the new equation

$$x(t) = A_1 e^{\alpha_1 t} + A_2 e^{\alpha_2 t} \tag{10}$$

where A_1, A_2, α_1, and α_2 are scalar constants, representing a system with two exponential components. As before, we use the Sum of Squared Error (SSE) as our loss function, leading to the following expression:

$$SSE_{2D} := \sum_i (A_1 e^{\alpha_1 t_i} + A_2 e^{\alpha_2 t_i} - x_i)^2 \tag{11}$$

Then we substitute $\beta_1 = e^{\alpha_1}$ and $\beta_2 = e^{\alpha_2}$, transforming the SSE_{2D} into a polynomial

$$SSE_{2D} = \sum_i (A_1 \beta_1^{t_i} + A_2 \beta_2^{t_i} - x_i)^2 \tag{12}$$

However, the progress in the multi-dimensional model has also emphasized the fact that the computational problems related to the introduction of more model parameters have become significant. Gröbner bases work well in the case of a few variables. However, as if you deal with a polynomial with a greater number of variables, you would quickly run into complexities and resource-intensiveness. This is because the complexity of Gröbner basis computation is exponential. As we are faced these computational constraints, we now turn to Homotopy Continuation as an alternative. This strategic move is motivated by the benefit of the technique's increased efficiency with a growing number of variables and its ability to do paralleled computations. Homotopy Continuation's way of path tracking, as opposed to the very lengthy computational cost for Gröbner basis, is a promising path of the minima exploration in the higher dimension model, while the prohibitive computational cost is avoided.

Initially, the first phase of this multi-dimensional investigation focused on an environment with absolutely zero noise and explored the effect of varying number of data-points on the number of positive roots, while holding all else constant. The experiment was conducted using the Python code in this Jupyter Notebook as well as this Julia code , and the results are presented table 6 below

Table 6: Number of Roots vs Number Datapoints Analysis

Number of Datapoints	**Mean Roots**	**Std Dev**
4	1	0.0
5	2	0.0
6	2	0.0
7	2	0.0
8	2	0.0
9	2	0.0
10	2	0.0
11	2	0.0
12	2	0.0
13	2	0.0
14	2	0.0
15	2	0.0
16	2	0.0

The observations seen in the table above substantiate the theoretical groundwork laid by Proposition 2.1 as it extended its applicability into a multi-dimensional case. The table consistently shows the presence of at least one positive root across all number of data-points tested. Moreover, the emergence of two roots becomes a recurrent pattern as we expand the number of data points. Furthermore, since we have been able to extend 2.1 into two-dimensional spaces, we hypothesize the following proposition.

Proposition 2.5. *Consider a dataset precisely conforming to an exponential growth model in $n \in \mathbb{N}$ dimensions, without any form of noise. In this scenario, the Sum of Squared Errors (SSE) function will have at least one single positive root, regardless of the number of data-points.*

In the next experiment, we dived further into the multi-dimensional case and aimed to understand the relationship between noise levels and the number of roots in higher dimensions. This experiment was conducted using a series of simulations, with each simulation incrementally increasing the noise level to observe its impact on the root landscape of the system. The Python portion of the experiment is documented in this Jupyter Notebook, while the root solving and analysis were performed with this Julia code. The results from these simulations are concisely summarized in Table 7.

Table 7: Number of Roots Various Noise Levels

Noise Level	Mean Roots	Std Dev	Max Roots	Min Roots
0	2.00	0.00	2.0	2.0
0.1	0.60	0.84	2.0	0.0
0.5	0.75	0.63	2.0	0.0
1	0.60	1.10	3.5	0.0
1.5	1.20	1.78	5.5	0.0
3	1.40	1.13	4.0	0.0
5	1.65	1.36	4.0	0.0
10	1.15	0.85	2.5	0.0
20	1.65	1.78	5.0	0.0
30	1.25	1.30	3.5	0.0
40	2.00	1.97	6.0	0.0
70	1.05	0.83	2.5	0.0
100	1.75	1.11	3.5	0.0
200	1.40	0.91	3.5	0.0
300	1.20	0.95	3.0	0.0
400	0.60	0.66	2.0	0.0
500	1.50	0.91	3.0	0.0

As we can see in the results summarized in Table 7, the mean number of roots shows a variable pattern as noise levels escalate. At a noise level of 0, we observe a consistent finding of 2 roots, affirming the previous noise-free experiments. However, as noise is introduced and gradually increased, we see fluctuations in the mean number of roots, alongside a broadening spread in their distribution as evidenced by the standard deviation. Particularly striking are the instances at higher noise levels where the maximum number of roots reaches up to 6, suggesting the presence of complicated loss function landscapes when we have significant noise levels. Moreover, the fluctuations in the minimum number of roots, dipping to 0 in several cases, further emphasizes the challenge of achieving reliable minima detection in noisy conditions. This observation aligns with intuitive expectations that noise complicates the identification of clear and stable solution paths. All in all, this last experiment exhibits a clear trend: as noise levels increase, the predictability and stability of the root detection process decrease. In other words, we get more chaotic and less predictable solution landscapes. These results are not too surprising as they fall in line with the previous one-dimensional case.

3 General Cases

In this section, we give some general propositions and proofs regarding number of roots and minima

Proposition 3.1. *Given two data points $(0, x_1)$ and $(1, x_2)$, the Sum of Squared Errors (SSE) has a single global minimum at $A = x_1$ and $\alpha = \log\left(\frac{x_2}{x_1}\right)$, where $\beta = e^{\alpha}$.*

Proof. The SSE for a model with parameters A and $\beta = e^{\alpha}$ given two data points is defined as:

$$SSE = (A - x_1)^2 + (A\beta - x_2)^2$$

Computing the gradient of SSE with respect to A and β, we obtain:

$$\frac{\partial SSE}{\partial A} = 2(A - x_1) + 2(A\beta - x_2)\beta$$
$$\frac{\partial SSE}{\partial \beta} = 2A^2\beta - 2Ax_2$$

Setting the gradient to zero for minimization:

$$2(A - x_1) + 2(A\beta - x_2)\beta = 0$$
$$2A^2\beta - 2Ax_2 = 0$$

Solving the system of equations yields $A = x_1$ and $\beta = \frac{x_2}{A} = \frac{x_2}{x_1}$. Recalling that $\beta = e^{\alpha}$, we find $\alpha = \log\left(\frac{x_2}{x_1}\right)$. To verify that this critical point is indeed a minimum, we simply check that the Hessian positive definite at this critical point.

Proposition 3.2. *Given 3 datapoints* $(0, x_1), (1, x_2), (2, x_3)$ *such that:*

- $x_1 < x_2 < x_3$
- $x_1, x_2, x_3 > 0$
- $x_1 \geq 1$
- $x_1 x_3 = x^2$

Then the SSE gradient has at most one positive roots, and so the SSE polynomial has at most one minima.

Proof. By definition, we have

$$SSE = (A\beta^0 - x_1)^2 + (A\beta^1 - x_2)^2 + (A\beta^2 - x_3)^2$$

We optimize the function by setting the gradient equal to zero

$$\nabla SSE = 0$$

Then we use Groebner basis to rewrite the above equation and obtain a polynomial in only one variable:

$$\begin{aligned} &b^6 x_2 x_3 + b^5(2x_1x_3 + x_2^2 - x_3^2) + b^4(3x_1x_2 - x_2x_3) \\ &+ b^3(2x_1^2 - 2x_3^2) + b^2(x_1x_2 - 3x_2x_3) \\ &+ b(x_1^2 - 2x_1x_3 - x_2^2) - x_1x_2 = 0 \end{aligned} \tag{13}$$

Now, if we apply the assumptions given in the proposition, then the polynomial abovein 13 has one and only one sign change, which by Descartes' rule of sign tells us that we has at most one positive root. Thus the SSE polynomial has at most one minima.

4 Conclusion

This thesis has investigated the complexities involved in estimating parameters for models described by differential equations, focusing specifically on the exponential growth models. Using Mathematics, Computer Science, and Computational methodologies, including the application of Gröbner bases and Homotopy Continuation, this work has illuminated the intricate landscapes of minima within the Sum of Squared Errors (SSE) framework. Below, we summarize some of the major findings and contributions of this research.

4.1 Summary of Major Findings

- **Impact of Noise on Minima Identification:** The analysis in this thesis has revealed a direct correlation between the noise level in the data and the number of minima in the SSE polynomial. This finding highlights the necessity for robust computational techniques when working on parameter estimation problems with noisy data.
- **Minimal Influence of Data Quantity on Minima Stability:** We have found that the number of datapoints, at least up to a point and within the tested model, has no effect on the number of minima of the loss function.
- **Multi-Dimensional Model Analysis:** By extending our study into two-dimensional models, we have been able to draw some valuable insights. We have found that there is a presence of at least one positive root across all evaluated scenarios, similar to the one dimensional case. This leads us to hypothesize that, for even higher dimensions, we will always have at least one positive root. Furthermore, in the two-dimensional model, we observed the emergence of two roots with the expansion of data points, highlights a nuanced interaction between model complexity and solution landscape. Multi-dimensional model also saw much more complex minima landscapes when noise was introduced.
- **Sensitivity to Non-Uniform Data Distribution:** For the one dimensional case, we found that the identification of minima is not significantly impacted by non-uniform distributions of time-stamps in the data collection process.

4.2 Contributions to Computer Science

This thesis has been able to contribute to the field of Computer Science through the following ways

- **New Computational Framework:** In this paper, we have fused Gröbner bases and Homotopy Continuation into parameter estimation. This has introduced a new computational framework capable of identifying the exact number and distribution of local minima for the loss function. This approach significantly improves the reliability of the parameter estimation process and offers a brand new toolkit that can be used by both researchers and practitioners alike.
- **Better Understanding of Optimization Landscapes:** Exploring how various factors influence the optimization landscape of the SSE function leads to a better knowledge and understanding of parameters estimation in differential equation models. These insight can be used for more competent optimization strategies, which are aimed at avoiding the risk of finding an 'non-optimal' local minima.
- **Foundational Work For Multi-Dimensional Cases:** The part of our analyses that extends to multi-dimensional models lays the foundations for future research endeavours.

4.3 Suggestions for Future Research

There are several directions in which future research can build upon the work in this report.

- **Higher-Dimensional Models**: Following in our footsteps and gain a deeper understanding of how these techniques can be used for higher dimensions.
- **Real-World Application and Validation:** The research done in this thesis has numerous practical applications. Hence, using it in a real-world setting will lead to further improvements, ad bridging the gap between the theoretical to the practical.

- **Integration with Machine Learning**: With machine learning on the rise at the time of writing this report, investigating the interaction between these algebraic approaches and machine learning could lead to more findings and novel methods.

5 References

[1] Saugata Basu, Richard Pollack, and Marie-Françoise Roy. *Algorithms in Real Algebraic Geometry*. Springer, 2 edition, 2006.

[2] B. Buchberger and M. Kauers. Groebner basis. *Scholarpedia*, 5(10):7763, 2010. revision #128998.

[3] Adarsh Prasad, Arun Sai Suggala, Sivaraman Balakrishnan, and Pradeep Ravikumar. Robust Estimation via Robust Gradient Estimation. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(3):601–627, 04 2020.

[4] Dustin Tran, Panos Toulis, and Edoardo M. Airoldi. Stochastic gradient descent methods for estimation with large data sets, 2015.

A Appendix

Statement of Academic Integrity Regarding Plagiarism

I, the undersigned.........Roudani Amine.........[family name, given name(s)], hereby certify on my honor that:

1. The results presented in this report are the product of my own work.

2. I am the original creator of this report.

3. I have not used sources or results from third parties without clearly stating thus and referencing them according to the recommended rules for providing bibliographic information.

Declaration to be copied below:

I hereby declare that this work contains no plagiarized material.

Date 18.03.2024

Signature