BACHELOR'S THESIS

(Arbeitstitel)

submitted to the

Institute of
Analysis and Scientific Computing
TU Wien

under the supervision of

Assistant Prof. Dr. Andreas Körner

by

Ida Hönigmann
Matriculation number: 12002348

Vienna, July 22, 2024

Acknowledgement

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Bachelorarbeit selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt bzw. die wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht habe.

Wien, am 22. Juli 2024

Ida Hönigmann

Contents

1 Introduction **1**

2 Optimization Algorithm **3**
- 2.0.1 Gradient Descent . . . 3
- 2.1 Least Square Problem Algorithms . . . 4
  - 2.1.1 Gauss–Newton Algorithm . . . 5
  - 2.1.2 Levenberg-Marquardt Algorithm . . . 6
  - 2.1.3 Trust Region Reflective Algorithm . . . 6
  - 2.1.4 Dogleg Algorithm with Rectangular Trust Regions . . . 6

3 T Cells, Calcium Concentration **9**
- 3.1 Components of a T Cell . . . 9
- 3.2 Activation . . . 10

4 Data **13**
- 4.1 Structure of Data . . . 13
- 4.2 How it was generated . . . 13
  - 4.2.1 Jurkat Cells, 5c.c7 primary mouse T cells and Fura-2 . . . 13
  - 4.2.2 Measuring Calcium Concentration . . . 14
  - 4.2.3 Processing . . . 15

5 Results **17**

6 Conclusion **19**

1 Introduction

Research questions:

- Which criteria can distinguish between unactivated, activated and pre-activated cells?
- Do different types of activated cells exists? How are they different?
- With which frequencies does the Calcium concentration repeat after activation?
- Is there a difference in frequencies between mouse and human cells?

2 Optimization Algorithm

An optimization problem is any problem where a function $f : X \to Y$ is given, and we search for the point $x \in X$ such that $f(x)$ is minimal or maximal. Obviously the minimum or maximum must not exist, as the example $f : (0, 1) \to \mathbb{R}, x \mapsto x$ demonstrates by not having either. Investigating conditions on X, Y and f such that a minimum or maximum exists is mathematically interesting. However, when implementing an optimization algorithm the true minimum or maximum can sometimes not be found even if it exists and is instead replaced by a sufficiently good approximation.

First we want to first think about some variations of the problem.

If a problem is stated to have additional conditions P the minimum or maximum must satisfy we can consider the subset $M := \{x \in X : P(x)\} \subseteq X$. By finding the minimum or maximum of $f : M \to Y$ a solution of the problem is solved. Once again such a minimum or maximum must not exist, even if we have that one is present in X.

In the example dealt with in this work we are given some data points $((x_k, y_k))_{k\in\{1,2,...,n\}}$ and want to find a close approximation in the form of a function $g(x, a_1, a_2, ..., a_m)$ where for every $a = (a_1, ..., a_m)$ we have a function $g_a(x) : \mathbb{R} \to \mathbb{R}, x \mapsto g(x, a_1, ..., a_m)$. Searching for a good approximation can be reformulated as searching for the minimum of $r(a) := \sum_{k=1}^{n} |g_a(x_k) - b_k|^2$ or any other error function. This form of optimization problem is called the Least Square Problem.

2.0.1 Gradient Descent

An iterative algorithm for finding the minimum of a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is gradient descent. As the name suggests it uses information of the gradient ∇f. Locally the negative gradient always points into the direction of greatest descent. The idea is to follow this direction for the next guess of the minimum. The pseudocode of this approach is given below.

Algorithm 1: Gradient Descent

input : $f : \mathbb{R}^n \to \mathbb{R}$ ... differentiable, $x_0 \in \mathbb{R}^n$
output: $x \in \mathbb{R}^n$

```
1  begin
2      for n = 0 to max_iterations do
3          if improvement is smaller than threshold then
4              break
5          end
6          set or calculate step size γ_n
7          x_{n+1} = x_n − γ_n ∇f(x_n)
8      end
9      x = x_n
10 end
```

If we consider a function with a local but not global minimum gradient descent might not converge to the optimum. An example of such a function can be seen in figure 2.1 along with the first values x_n of gradient descent for a starting value not converging to the global minimum.

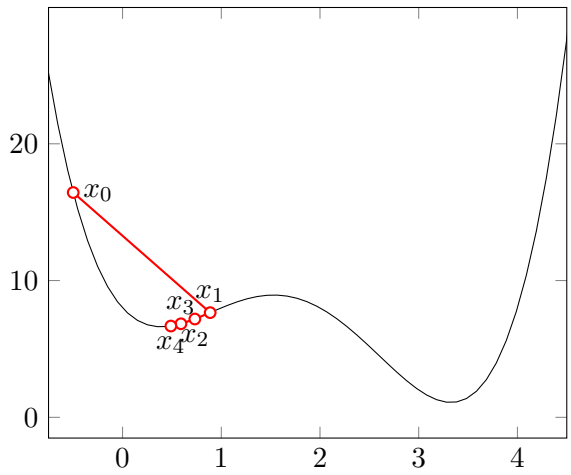

Figure 2.1: The function has two local minimums. For this starting value and step size gradient descent approaches the local, but not global minimum.

Improvements can be made by choosing good step sizes, starting value or by starting with different values and comparing the results.

2.1 Least Square Problem Algorithms

We now focus on the Least Square Problem and give an introduction into various algorithms used.

Formally we are given a residual function $r_f(x)$ which tells us whether a function f is a good approximation at the point x. We therefore want to find a way to minimize $||r(x)||^2$.

2.1.1 Gauss–Newton Algorithm

The idea behind this algorithm is that it is easy to find the intersection with zero of a linear function. If we linearize $r(x)$ locally we can approximate the root by finding it of the linear approximating function. This is demonstrated in figure 2.2.

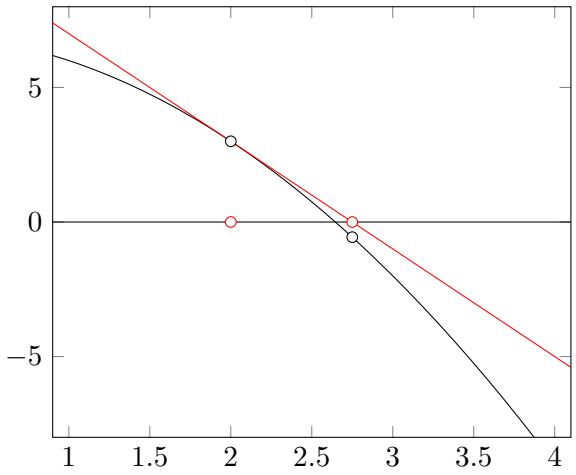

Figure 2.2: By approximating the black function by a line an approximation of the root has been found.

Iterating this step of linear approximating gives us the Gauss-Newton Method. In figure 2.3 we can see that indeed x_n seems to converge towards the root of the function.

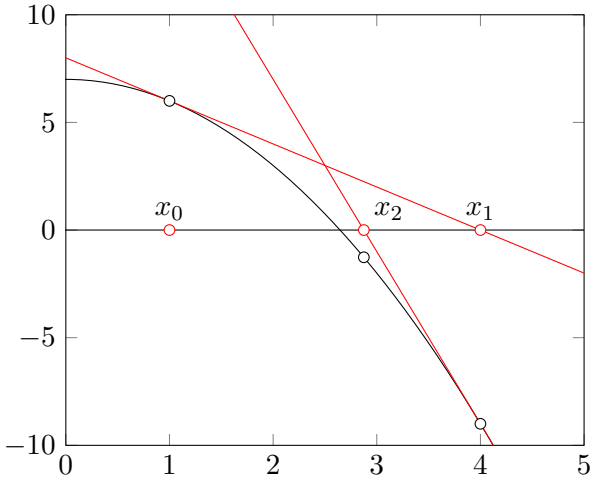

Figure 2.3: Iteratively applying linear approximation gives the Gauss-Newton Method for approximating the root.

Define Dr as the Jacobian matrix $\left(\frac{\partial r_i}{\partial x_j}\right)_{ij}$. Using Taylor's theorem we get the linear approximation

$$r(x) = r(a) + Dr(a)(x-a) + h(x)(x-a) \approx r(a) + Dr(a)(x-a) \text{ with } \lim_{x \to a} h(x) = 0.$$

Rewriting this as $r(x) \approx Ax - b$ where $A := Dr(a)$ and $b := Dr(a)a - r(a)$ gives us the algorithm for this method. As $Dr \in \mathbb{R}^{n \times m}$ we solve $Dr^T Drx = Dr^T b$ in order to get a system with square matrix. If $n = m$ we can skip this step and get the so-called Newton algorithm as a variant.

Algorithm 2: Gauss-Newton

input : $r : \mathbb{R}^n \to \mathbb{R}^m$ ... differentiable, $x_0 \in \mathbb{R}^n$
output: $x \in \mathbb{R}^n$

```
1  begin
2      for n = 0 to max_iterations do
3          if x_n close enough to zero then
4              break
5          end
6          Calculate A_n := Dr(x_n)
7          Calculate b_n := A_n x_n − r(x_n)
8          Solve A_n^T A_n x_{n+1} = A_n^T b_n
9      end
10     x = x_n
11 end
```

Gauss-Newton is guaranteed to find a local minimum x if r is twice continuously differentiable in an open convex set including x, Dr has a full rank and the initial value is close enough to x.

For the example demonstrated in figure 2.4 we can see that choosing a particular starting value leads to a loop in which only two points are explored as possible roots. More extreme examples exists in which Gauss-Newton gets increasingly further away from the root, due to an increasingly flat incline the further we get from the root. One example of such a function can be seen in figure 2.5.

Both problems, the starting value being too far from the root and the Dr not having full rank, can be combated using the technique of dampening. Instead of moving the new guess all the way to the root of the linear approximation we only move part of the way. How much can be determined by a dampening factor λ_n or a constant λ.

2.1.2 Levenberg-Marquardt Algorithm

2.1.3 Trust Region Reflective Algorithm

2.1.4 Dogleg Algorithm with Rectangular Trust Regions

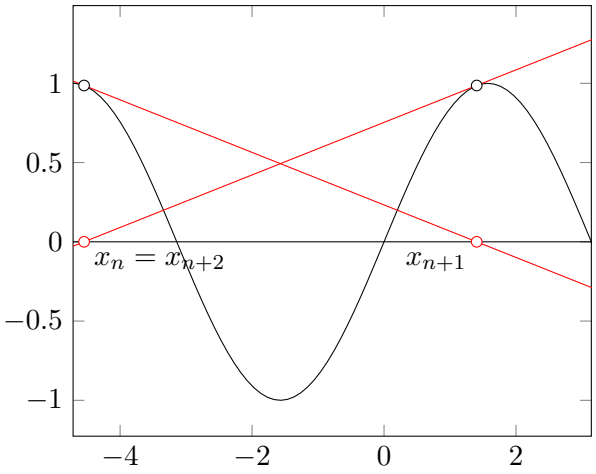

Figure 2.4: For a poor choice of starting values Gauss-Newton can never find the root of the function $\sin(x)$.

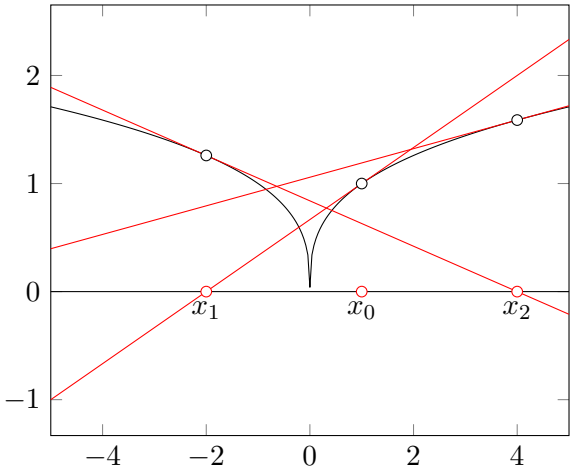

Figure 2.5: Finding the root of the function $\sqrt[3]{|x|}$ using Gauss-Newton is only possible if the starting value x_0 is chosen as 0, which is the root. For any other value we have that the guess gets further and further away. Indeed for any x_n we have $x_{n+1} = -2x_n$.

3 T Cells, Calcium Concentration

Lymphocytes form a key component of the immune system. T cells are a type of lymphocyte and are responsible for responding to viruses, fungi, allergens and tumours. Different subtypes of t cells exist, that perform various responsibilities. They are transported throughout the body via the lymphatic system and blood.[**Kumar2018**]

Precursor cells are formed in the bone marrow. Once they are transported to the thymus they undergo maturation and selection to become t cells. Each cell forms receptors, called t cell receptors (TCR), that respond to one particular out of many (10^6–10^9) possible short pieces of proteins, called peptides. These peptides are attached to the major histocompatibility complex (MHC) present on antigens and antigen presenting cells (APC). Important aspects of the selection are ensuring that the t cells react to foreign peptides, but not to those present on the body's own cells.[**Ashby2024**]

In positive selection cells in the thymus present peptides on their MHC. If a t cell is unable to bind, it will undergo apoptosis, a type of cell death. T cells which were able to bind receive survival signals. Negative selection verifies that t cells will not attack the body's own cells. This is done by only selecting t cells which only bind moderately to the peptides presented, as a strong bond suggests that these t cells would have a high likelihood of being reactive to own cells.[**Hagel2018**] If a t cell passed both the positive and negative selection it is transported to the periphery.

There are multiple types of peripheral t cells. Native t cells respond to new antigens. Cytotoxic t cells kill cells which present peptides on their MHC compatible with the t cells TCR. Helper T cells activate other parts of the immune response. Memory t cells shorten the reaction time when the same antigen is encountered again at a later point in time. Suppressor t cells moderate the immune response.[**Ganong1997**]

3.1 Components of a T Cell

T cell components relevant in activation and subsequent changes in intracellular Ca^{2+} are listed below and schematically shown in figure 3.1.

- **T cell receptor (TCR):** Receptor on the cell surface that can recognize peptides. By the simultaneous triggering of the TCR and co-stimulator signalling is induced that leads to activation.
- **Co-stimulator:** A stimulation of co-stimulatory molecules is necessary in order for signalling to occur as part of activation.
- **Endoplasmic reticulum (ER):** A series of connected sacs in the cytoplasm that is attached to the nucleus. Important functions are folding, modification and transportation of proteins.[**Rogers2024**]

- **Ca^{2+} permeable ion channel on the ER:** There are several Ca^{2+} channels present on the ER. Some receptors are responsible for releasing Ca^{2+} into the cytoplasm, when the intracellular Ca^{2+} concentration is low. [**Schwarz2016**]
- **Ca^{2+} storage in the ER:** Ca^{2+} is stored in the ER and can be released by Ca^{2+} permeable ion channels on the ER.
- **Cytoplasm:** The semi-fluid substance enclosed in the plasm membrane. It contains organelles, ions, proteins and molecules.
- **Stromal interaction molecule (STIM):** If the Ca^{2+} storage in the ER is depleted STIM proteins cluster where the ER is in the vicinity of the plasm membrane and assembles CRAC, which then leads to uptake in extracellular Ca^{2+}. [**Schwarz2016**]
- **Plasm membrane:** A semipermeable structure forming the wall of the cell made up of lipids and proteins. Ion channels and transport proteins allow certain substances to move through.[**Ganong2012**]
- **Ca^{2+} release activated Ca^{2+} channel (CRAC):** Opened after a decrease in ER stored Ca^{2+} is sensed by STIM, these channels intake Ca^{2+} from outside the cell.[**Stathopulos2013**]
- **Cytoskeleton:** A system of fibres within the cell, that allows it to change shape and move.[**Ganong2012**]
- **Nucleus:** An organelle that stores most of the DNA, controls cell growth and cell division. A double membrane separates it from the cytoplasm.[**cooper2022**]

Relevant components of APC are the

- **Major histocompatibility complex (MHC)**, which can present peptides, and the
- **Co-stimulator**, which can form a bond with the co-stimulator on a t cell.

Both are present on the surface of the APC.

3.2 Activation

Activation is necessary for t cells to divide and perform their functions.[**Ganong1997**]

When a native t cell encounters a peptide on an APC that is compatible, a bond is formed between the TCR on the t cell and the peptide-MHC complex on the APC. This recognition can be triggered by less than ten molecules of foreign substance and is therefore described as near perfect. Sufficiently long contact is necessary between the APC and the t cell in order for the t cell to activate. The role of contact time in t cell activation is modelled by Morgan et al..[**morgan2023**].

The presence of co-stimulatory molecules is needed for proper activation. The bond between the co-stimulatory molecules on the t cell and APC plays a role in signalling. Ca^{2+} signals play a vital part in t cell activation.

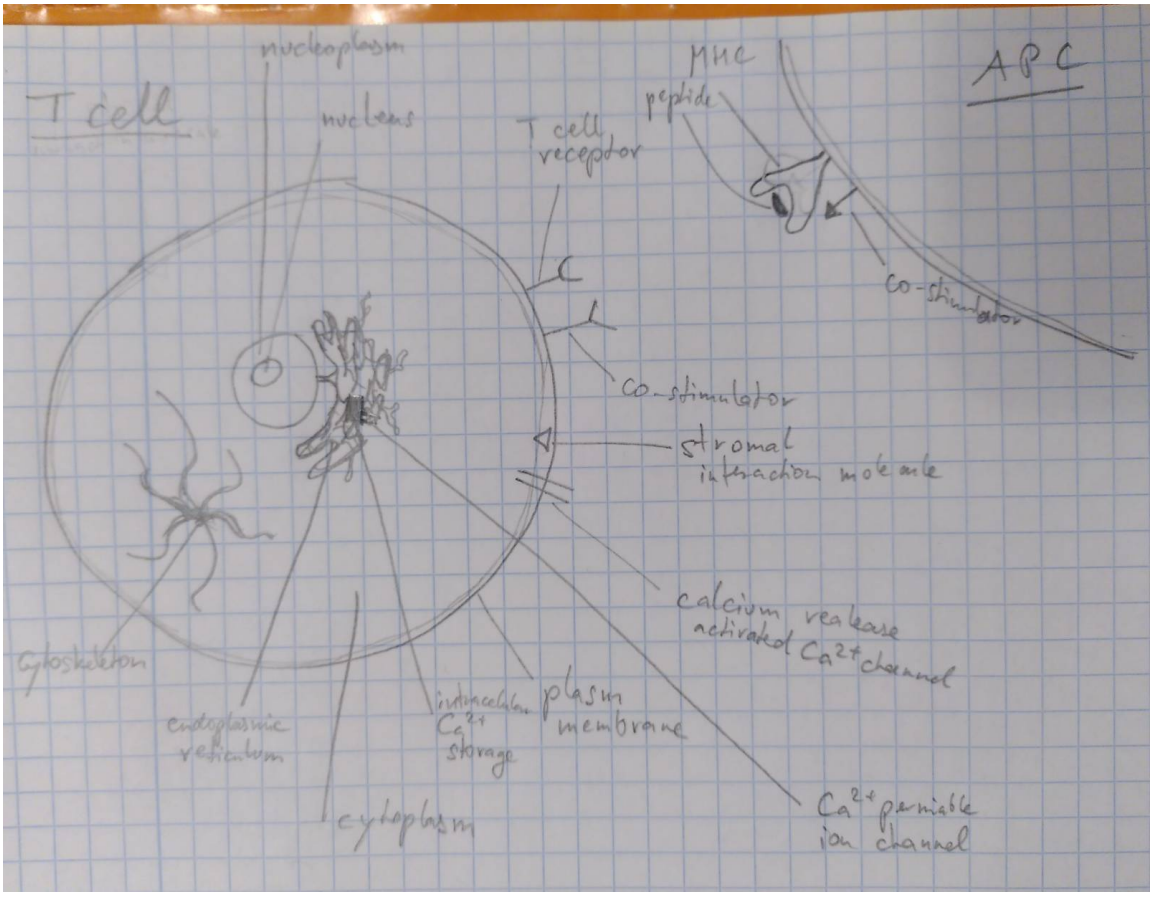

Figure 3.1: Schematic view of a t cell and antigen presenting cell, with all relevant components.

An increase of Ca^{2+} in t cells during activation is caused by the stimulation of Ca^{2+} permeable ion channel receptors on the ER membrane. Ca^{2+} is released from the ER into the cytoplasm. Additionally, this decrease in Ca^{2+} is sensed by STIM, which leads to an influx of Ca^{2+} through plasma membrane CRAC channels.[**smith2009**]

As the intracellular Ca^{2+} concentration is dependent on the interaction between Ca^{2+} sources and sinks, a variety of different forms in Ca^{2+} concentration have been observed. Examples are infrequent spikes, sustained oscillations and plateaus.[**Lewis2001**]

Intercellular Ca^{2+} increase together with other signals lead to a redistribution of receptors, signalling molecules and organelles.[**joseph2014**]

4 Data

From section 3.2, we gather that analysing the intracellular Ca^{2+} concentration gives us good insight in whether and when a cell activates. Additionally, it can be measured relatively easily by the method described in this chapter.

4.1 Structure of Data

The data matrix has a row per tracked cell and frame. The information stored for each cell and frame combination is described in detail in table 4.1.

Name	Data Type	Description
x	float64	Position of cell in pixels along the horizontal axis
y	float64	Position of cell in pixels along the vertical axis
frame	int32	Number of frame, with frame rate of 1 frame per second
mass short	float64	Brightness of cell in 340nm channel
bg short	float64	Background in 340nm channel
mass long	float64	Brightness of cell in 380nm channel
bg long	float64	Background in 380nm channel
ratio	float64	Calculated as mass short divided by mass long
particle	int32	Identification for each particle

Table 4.1: Description and data type of all columns present in the data matrix.

One recording can have between 500 and 10000 cells and is between 700 and 1000 frames long, which corresponds to between about 11 and 17 minutes. The ratio is typically between 0 and 5.

4.2 How it was generated

Four recordings where generated, with two each from human and mouse cells. Per cell type a positive and negative control was measured. In a positive control the conditions are such, that in theory every cell should activate, while in negative control the conditions are such, that none should activate. Due to stress on the cells and other factors a few cells will not follow theory.

4.2.1 Jurkat Cells, 5c.c7 primary mouse T cells and Fura-2

The prototypical cell line to study T cell signalling is the Jurkat cell line.[**morgan2023**] It was obtained from the blood of a boy with T cell leukaemia.[**schneider1977**] Different

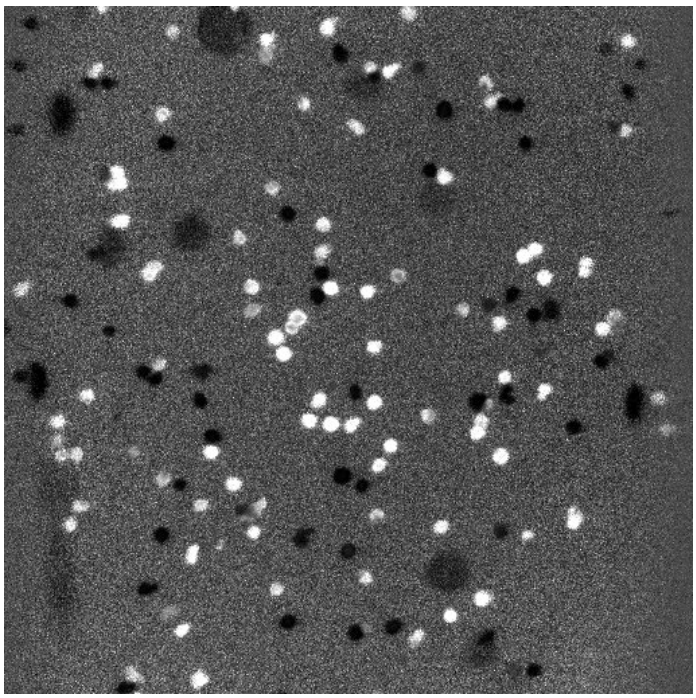

Figure 4.1: Single frame showing the ratio of the 340nm and 380nm images from a recording of human Jurkat cells. Activated cells appear lighter, unactivated cells darker than the background. Big dark circles are out of focus cells that have not yet settled on to the plate.

cell lines within the Jurkat family are described by Abraham and Weiss.[**abraham2004**] They provide a timeline of discoveries linked to Jurkat cells and t cell receptor signalling.

Another type of T cells used in signalling studies are gathered from mice. [Additional information]

In order to be able to measure the intracellular Ca^{2+} concentration of cells they can be labelled with Fura-2. This method provides a way to record the Ca^{2+} concentration of multiple cells over a time period.[**martinez2017**] Challenges encountered when using Fura-2 on certain cell types are described by Roe, Lemasters and Herman along with their respective solutions.[**roe1990**]

4.2.2 Measuring Calcium Concentration

After the cells have been labelled with Fura-2, a recording of up to 15 to 20 minute can be generated. To achieve this the cells and stimulant are photographed at both 340nm and 380nm wavelength once per second. The resolution of the images are 1.6um per pixel. By calculating the ratio of the two images at each point the Ca^{2+} concentration can be observed. An exemplary resulting image showing the ratio is shown in figure 4.1. The T cells appear a lighter shade than the background when activated and darker when not activated.

To activate the cells in the duration of the recording they are transferred to a plate

covered with replicas of the MHC-peptide complex normally present on APCs. This plate is then recorded as described above. For a negative control the plate is not covered with peptides, while for the positive control the peptide covering on the plate is very dense. Recordings of different densities in peptides lead to activation of a percentage of t cells.

4.2.3 Processing

To track single t cells moving around during the video the sum of the 340nm and 380nm image of each second is calculated. This image provides the basis for separating t cells from the background. On this image all t cells will appear similarly light in colour. Therefore, it is used to track the movement of cells. Each cell is numbered, such that the same cell will have the same number during the video. For some cells the trajectory tracking is not perfect, resulting in a split of the numbering into multiple numbers for the same cell. The position and shade during both 340nm and 380nm as well as the ratio of each particle and each frame is then recorded into the data structure used in this work. The first roughly 50 frames at the start of the recording are discarded due to the video being out of focus. Additionally, cells only appearing in fewer than 300 frames are discarded as they most likely represent trajectories incorrectly tracked or split. The resulting data is then stored in a matrix structured as described in table 4.1.

5 Results

6 Conclusion