# A collection of modelling problems carried out in the academic year 1992-1993 by the ECMI-students at the Eindhoven University of Technology

Technische
Universiteit
Eindhoven

# Opleiding Wiskunde voor de Industrie Eindhoven

## Student Report 94-04

# A collection of modelling problems 1992 - 1993

Hans Bonekamp
Marijke van der Burg
Richard du Croo de Jongh
Mariken Everdij
Olivier le Flahec
Klaas Friso
Alexander Franz
Jeroen Gaarthuis
Marcel Grob
René Haarman
Jozef Haslinger
Dries Hegen
Anh Hoang
Koen de Jong
Jurgen Kok
Christa König
Robert Linders
Niels Lous
Astid Mirle
Adri Mourits
Gerard Pruis
Maria Anna Stadler
Serghey Tiourine
Berno Wibbels

ECMI

Den Dolech 2
Postbus 513
5600 MB Eindhoven

December 1993

A collection of Modelling Problems carried out in the academic year 1992 - 1993 by the ECMI-students at the Eindhoven University of Technology

Hans Bonekamp
Marijke van der Burg
Richard du Croo de Jongh
Mariken Everdij
Olivier le Flahec
Klaas Friso
Alexander Franz
Jeroen Gaarthuis
Marcel Grob
René Haarman
Josef Haslinger
Dries Hegen
Anh Hoang
Koen de Jong
Jurgen Kok
Christa König
Robert Linders
Niels Lous
Astrid Mirle
Adri Mourits
Gerard Pruis
Maria Anna Stadler
Serghey Tiourine
Berno Wibbels

# Table of contents

# MARKING THE STARTING SPOTS ON AN IMPROVED ATHLETICS TRACK

# Table of Contents

# Abstract

Athletics tracks, as we know them now, consist of straight lines and half circles. The length of such tracks is easy to calculate. Knowing the length of a track, the starting spots for the various distances are easily marked. A disadvantage of ordinary athletics tracks is that the curvature is not a continuous function of the track distance. This means that an athlete suddenly has to start and stop his turn. New tracks can be developed in which the curvature is a continuous function of the track distance. These tracks would show an improvement compared with the old tracks. However, the length of these tracks is not easy to determine.

The purpose of this report is to calculate the length of a class of specific curves. These curves consist of parts of circles and of clothoid splines. New tracks based on these curves, should be faster than the regular tracks. The starting spots of the different distances are marked on the new tracks.



# Acknowledgements

The task reported here was undertaken in the context of the modelling practical study of the ECMI postgraduate programme "Mathematics for Industry" at the Eindhoven University of Technology in the Netherlands. The goal of this practical study is to give quick and useful answers to various questions that involve mathematics. We were assisted in this practical study by Dr Ir S J L van Eijndhoven.

This report is used as thesis work in the course Technical Writing and Editing given by Dr C A Hawkins and Mvr M Andrews, whom the authors wish to thank for their comments on general writing technics.

The authors also thank Mr van der Sommen (Sports union of the Eindhoven University of Technology) for his help, and Mr Moormans (Netherlands Athletics Union) for his useful information on athletics tracks.

# 1 Introduction

## 1.1 Rules for building athletics tracks

An official athletics track has a length of 400 m. This length is measured 30 cm from the inside of the track. A track can have at most eight lanes. The usual number of lanes is either six or eight. The width of one lane can be somewhere between 1.22 m and 1.25 m. The first lane from inside is 10 cm wider. There are rules for the maximum slope of the track in the turns and for the thickness of the lines. Strange enough, the slope can either be to the inside of the track or to the outside. They both occur in practice.

The athletic distances start in the different lanes at different positions on the track. The athletes run in anti-clockwise direction. There is only one finish line for all lanes. This line is marked at the end of one straight. The straight containing the finish line is called the home straight while the other straight is called the back straight. For some reason, according to some literature, the home straight should be West-East directed.

## 1.2 Freedom in choice for the actual track shape

Although there are several rules for the width, the length and the slope of a track, there is no rule for the actual shape. Tracks are built with different shapes. The most common choice is a track consisting of two straight lines and two half circle curves (see figure 1.1), henceforward indicated as regular tracks. The radius of the curves can still be chosen. In the English literature, they recommend a radius larger than 32 m. A radius of 40 m is normal in the Netherlands so that a football field fits inside the track. In countries where athletics is more important than in the Netherlands, a radius of 36.5 m is more common. In the area enclosed by these tracks, other athletic disciplines can be performed.

Some tracks have curves that consist of circle parts with two different radii (see figure 1.2). The curve starts and ends with a circle part with small radius (e.g., 24 m) and has a circle part with large radius (e.g., 48 m) in the middle. The advantage of these tracks is that the dimensions of the track are smaller compared with those described above, although the straight parts are longer.

In this report we will discuss an athletics track with curves that consist of circle parts and clothoidal curves. Curves of this kind are used in railway engineering and in road construction. Theoretically, these curves are in some way better than the original tracks. In the original tracks, the curvature is not a continuous function of the track distance. The curvature is equal to zero on a straight and equal to a constant different from zero in a curve. In the new tracks we will connect the straight and the circle parts with clothoid splines. These splines have linearly growing curvature over the track distance.

It is believed that athletes can run the clothoidal tracks faster than the original tracks. Some research into this kind of tracks has been done in Switzerland. It is much more difficult to mark the exact starting spots on the clothoidal tracks than on the circular curved tracks.

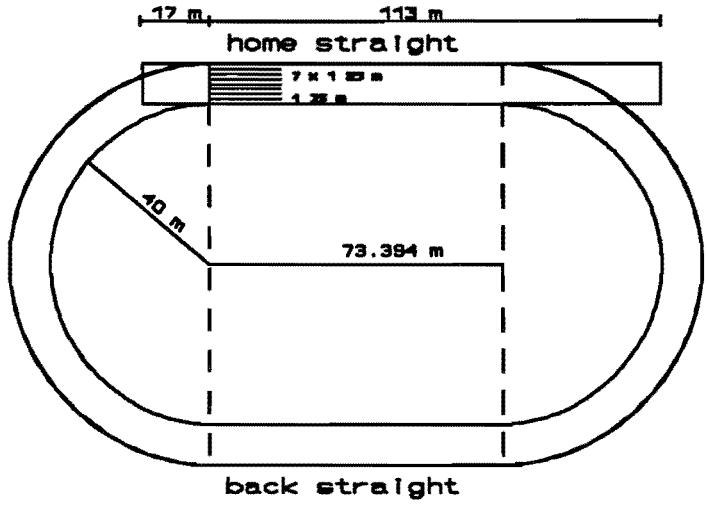



**Figure 1.1** *A regular Dutch athletics track.*

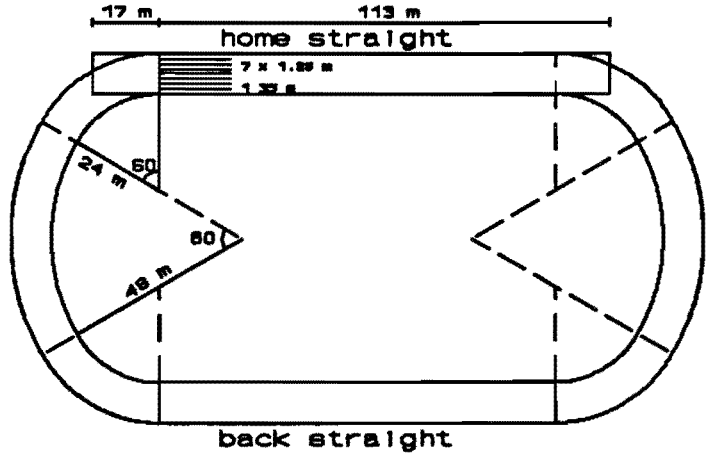



**Figure 1.2** *An athletics track with curves, consisting of different circle parts.*

### 1.3 Contents of this report

The basic questions we will consider in this report are formulated in chapter 2. The assumptions we made are postulated in chapter 3. General answers on the questions are given in chapter 4. The mathematical modelling of the problem is given in chapter 5. Some specific (numerical) results are given in chapter 6. Finally, in chapter 7, the conclusions and recommendations are considered.

# 2 Problem description

The assignment given was to construct an athletics track and to point out where the starting spots for five athletic distances should be marked. These distances are respectively the 100, 200, 400, 800 and 1500 m. In this chapter we will put the problem in a more specific way.

### 2.1 Regular athletics tracks

A regular athletics track consists of two straights and two half circle curves. The radius of the curves can be chosen freely within certain limits. For tracks with a radius of 40 m and of 36.5 m, we will mark the starting spots for the distances mentioned above.

### 2.2 An athletics track with continuous curvature

The main goal of this report is to construct an athletics track which has continuous curvature. We will use clothoids to make a connection between the straights and the circle parts. A good choice for the length of the clothoidal parts and the curvature of the circle part must be made. Finally, the starting spots for the distances mentioned above are again to be marked.

# 3 Assumptions

We will consider a track with eight lanes. Note that the starting spots on a track with six lanes are the same as the starting spots of the first six lanes from inside of an eight lane track. The width of the lanes is 1.25 m with exception of the first lane, which is 1.35 m wide. The length of lane two to eight is measured 20 cm from the inner side of the lane. The length of lane one is measured 30 cm from inside. The slope of the track will not be considered.

## 3.1 Athletic distances

We will only mark the starting spots for the 200 m and the 400 m. An explanation for not considering the other distances is given in this paragraph.

### 3.1.1 100 meters

The 100 m sprint is always held on one straight line. Because the straights are always shorter than 130 m (3 m start space, 110 m hurdles and 17 m run out space), the home straight is extended. The starting spots for the 100 m can easily be marked on these straight lanes.

### 3.1.2 800 meters

The starting spots for the 800 m are nearly the same as for the 200 m, shifted over a half track. In fact, the 800 m race consists of two parts. In the first part the athletes keep their own lane. In the last one-and-a-half lap, the athletes all run in the first lane. This part is exactly 600 m long. The change from the first part to the second part is made on the back straight. Here the athletes in the lanes two to eight run diagonally from their lane to lane one. A correction for the extra distance the athletes have to run is made on the starting spots for the 200 m. The following example shows the magnitude of this correction term. Consider a circular curved track with a radius of 40 m. The length of the straights is then 73.394 m. The correction terms for lanes one to eight are given in table 3.1.

| lane number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| correction term | 0.0 | 1.1 | 4.3 | 9.6 | 17.0 | 26.6 | 38.2 | 52.0 |

**table 3.1** *Correction in centimeters on the 200 m starting spots in order to mark the 800 m for a circular curved track with 40 m radius.*

### 3.1.3 1500 meters

Athletes running the 1500 m all begin their race at the same line. They all run in the first lane. Therefore, the start line has to be 300 m before the finish line. It is only necessary to mark the 300 m point in the first lane.

# 4 Solutions

The results in this chapter are given in the special coordinate system $(x,r,\varphi)$ which is defined in paragraph 5.5. We recall here that $x$ gives the polar origin, $r$ the radius and $\varphi$ an angle compared with the normal to the $x$-axis, taken positive in anti-clockwise direction, such that the negative $x$-direction corresponds with an angle $\pi/2$.

## 4.1 Starting spots on regular athletics tracks

It is easy to calculate the starting spots on regular athletics tracks for all athletic distances if the radius of the curves and the width of the lanes are given. Tables 4.1 and 4.2 show the geometric measures that point out the starting spots for the 200 m and the 400 m on regular tracks with radius of 36.5 m and 40 m and a lane width of 1.25 m.

## 4.2 Starting spots on a clothoidal athletics track

Although it is more work compared with regular tracks, it is not difficult to calculate the starting spots on a clothoidal athletics track. To mark the spots, however, is not very easy. The clearest way to give the co-ordinates is by expressing them in our self-defined co-ordinate system $(x,r,\varphi)$. In all cases two Fresnel integrals must be calculated numerically. In tabel 4.3 the starting spots for the 200 meter and the 400 meter for a clothoidal track with radius $R$ = 36.5 meter and parameter $B$ = 18.25 are given in the co-ordinate system $(x,r,\varphi)$.

| lane | 200 meter | | | 400 meter | | |
|---|---|---|---|---|---|---|
| | $x$ | $r$ | $\varphi$ | $x$ | $r$ | $\varphi$ |
| 1 | 42.195 | 36.800 | 3.142 | -42.195 | 36.800 | 0.000 |
| 2 | 42.195 | 38.050 | 3.245 | -42.195 | 38.050 | 0.206 |
| 3 | 42.195 | 39.300 | 3.341 | -42.195 | 39.300 | 0.400 |
| 4 | 42.195 | 40.550 | 3.432 | -42.195 | 40.550 | 0.581 |
| 5 | 42.195 | 41.800 | 3.517 | -42.195 | 41.800 | 0.752 |
| 6 | 42.195 | 43.050 | 3.598 | -42.195 | 43.050 | 0.912 |
| 7 | 42.195 | 44.300 | 3.673 | -42.195 | 44.300 | 1.064 |
| 8 | 42.195 | 45.550 | 3.745 | -42.195 | 45.550 | 1.207 |

**table 4.1** *Co-ordinates x and r in meters and φ for the 200 and the 400 meters on a regular track with radius R = 36.5 meter.*

| lane | 200 meter | | | 400 meter | | |
|---|---|---|---|---|---|---|
| | $x$ | $r$ | $\varphi$ | $x$ | $r$ | $\varphi$ |
| 1 | 36.697 | 40.300 | 3.142 | -36.697 | 40.300 | 0.000 |
| 2 | 36.697 | 41.550 | 3.236 | -36.697 | 41.550 | 0.189 |
| 3 | 36.697 | 42.800 | 3.325 | -36.697 | 42.800 | 0.367 |
| 4 | 36.697 | 44.050 | 3.409 | -36.697 | 44.050 | 0.535 |
| 5 | 36.697 | 45.300 | 3.488 | -36.697 | 45.300 | 0.694 |
| 6 | 36.697 | 46.550 | 3.563 | -36.697 | 46.550 | 0.844 |
| 7 | 36.697 | 47.800 | 3.635 | -36.697 | 47.800 | 0.986 |
| 8 | 36.697 | 49.050 | 3.702 | -36.697 | 49.050 | 1.121 |

**table 4.2** *Co-ordinates x and r in meters and φ for the 200 and the 400 meters on a regular track with radius R = 40 meter.*

| lane | 200 meter | | | 400 meter | | |
|---|---|---|---|---|---|---|
| | $x$ | $r$ | $\varphi$ | $x$ | $r$ | $\varphi$ |
| 1 | 27.861 | 37.733 | 3.142 | -27.861 | 37.733 | 0.000 |
| 2 | 31.488 | 38.975 | 3.149 | -35.454 | 38.893 | 0.039 |
| 3 | 34.572 | 40.173 | 3.170 | -40.143 | 39.744 | 0.156 |
| 4 | 37.017 | 41.314 | 3.205 | -42.038 | 40.579 | 0.337 |
| 5 | 38.925 | 42.403 | 3.251 | -42.121 | 41.800 | 0.409 |
| 6 | 40.327 | 43.465 | 3.306 | -42.121 | 43.050 | 0.579 |
| 7 | 41.276 | 44.533 | 3.369 | -42.121 | 44.300 | 0.740 |
| 8 | 41.852 | 45.636 | 3.436 | -42.121 | 45.550 | 0.892 |

**table 4.3** *Co-ordinates x and r in meters and $\varphi$ for the 200 and the 400 meters on a clothoidal track with radius $R = 36.5$ meter and with parameter $B = 18.25$.*

# 5 Mathematical model

In this chapter, we will calculate the location of the starting spots on two different athletics tracks. The difficulty in marking the starting spots is to find the length of each lane. Moreover, the length of each part of a lane is to be known. In paragraphs 5.2 and 5.4 these lengths will be calculated for the two different tracks. First we will consider the regular tracks, then we will consider the clothoidal tracks. The transformation of mathematical results into geometrical measures is given in paragraph 5.5.

## 5.1 Introduction

The length of a lane is calculated by measuring each part of the track, starting at the finish line, going in clockwise direction. The lane is divided into different parts. The first division for instance is made by splitting the lane in two straights and two turns. The length of each part is calculated.

## 5.2 Regular athletics tracks

In this paragraph we will mark the starting spots on the regular circular curved athletics tracks. Henceforward we suppose that an athletic distance $D$, a radius $R$ of the curves and a width $w$ of the lanes are given.

### 5.2.1 Track length

First we divide the track into four parts: two straights and two curves. Because the curves are half circles, the length of the curves are easily found. The curve length of lane one is calculated as

$$s_1 = (R + 0.30)\pi,$$

and those of the other lanes as

$$s_i = s_1 + (i-1)w\pi, \qquad i = 2,3,..8.$$

Note that the curve length of lane one is calculated 30 cm from the inside of the track which has radius $R$. Because the length of the first lane is always 400 m, the length of one straight $L$ must be

$$L = 200 - s_1.$$

The length of both the straights and the curves of each lane is known at this point. Now we can locate the starting spots.

### 5.2.2 Starting spots

We subtract as many half track lengths as possible from the distance $D$. We also calculate the rest term $e_i$.

$$k_i = D \text{ div } (L + s_i),$$
$$e_i = D - (L + s_i)k_i, \qquad i = 1,2,..8.$$

The rest term $e_i$ is a number that is larger or equal to zero, but smaller than half the track length of lane $i$. A distinction is made between two different cases.

If the rest term is smaller than $L$ than the starting spot for that lane is located on a straight. The term $e_i$ is the distance between the starting spot in lane $i$ and the end of the straight. If $k_i$ is even, the spot is located on the home straight, otherwise the spot is located on the back straight.

If the rest term $e_i$ is larger than $L$, the starting spot is located in a curve. The easiest way to indicate the starting spot is by giving the angle between the normal to the track at this spot and the positive vertical axis. This angle is equal to

$$\varphi_i = \left(2 - \left[k_i\right]_2 - \frac{e_i - L}{s_i}\right)\pi,$$

where $[\ ]_2$ denotes modulo 2. The angle grows in anti-clockwise direction.

## 5.3 Continuous curvature

A regular track has discontinuous curvature over the track distance. Figure 5.1 shows the curvature in and near a curve of a regular track. We believe that by making the curvature of the track continuous the track should be faster.

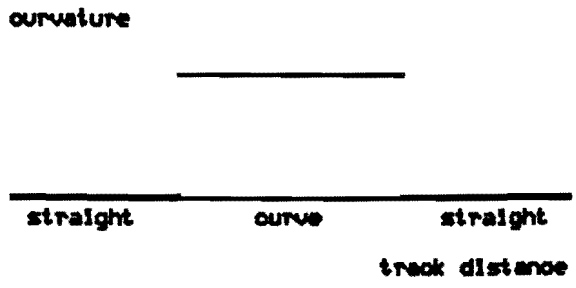



**Figure 5.1** *Curvature of a circular curved track in the neighbourhood of a curve as function of the track distance.*

The surface under the graph is per definition equal to the total turn in one curve and this must be equal to $\pi$. Furthermore the maximum curvature must not be to large and the total curve should not occupy to much track distance. With these restrictions, several tracks with continuous curvature can be constructed. We make the curvature continuous by inserting in the regular track two curves with linearly growing curvature (see figure 5.2). A curve that has linearly growing curvature is called a clothoid.

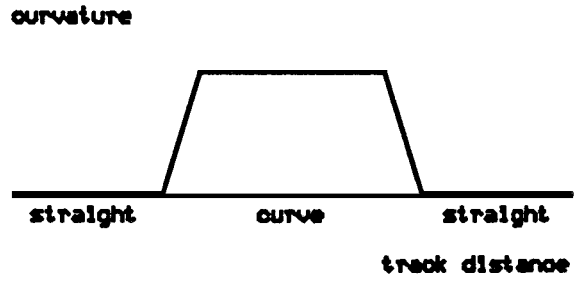



**Figure 5.2** *Curvature of an improved track in the neighbourhood of a curve as function of the track distance.*

A choice for the maximum curvature and for the slope of the ascending and descending part of the curve is to be made. Given these values and given that hte surface under the graph must be equal to $\pi$, the curve is determined. The length of the straights follows from the fact that the first lane must be 400 meters long. In this article we will choose a minimum curve radius of 36.5 m. The choice for the slope can be left free for optimization. We will consider only one slope.

### 5.3.1 Clothoid splines

According to Meek and Thomas [1], clothoid splines are defined parametrically by

$$f_B(t) : \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \pi B \begin{pmatrix} C(t) \\ S(t) \end{pmatrix},$$

where the scalings factor $\pi B$ is positive, the parameter $t$ is nonnegative and where $C(t)$ and $S(t)$ are the Fresnel integrals

$$C(t) = \int_0^t \cos\tfrac{1}{2}\pi u^2 \, du, \qquad S(t) = \int_0^t \sin\tfrac{1}{2}\pi u^2 \, du.$$

A graph of a clothoid spline is shown in figure 5.3.

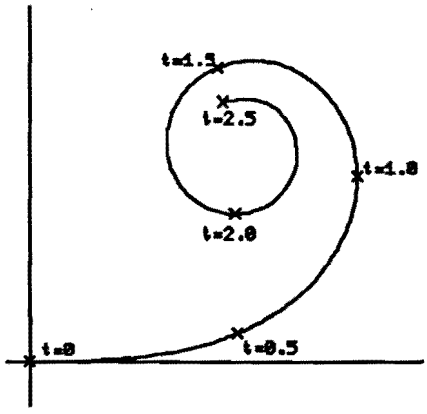



**Figure 5.3** *A part of a clothoid.*

Some characterisitics of the clothoid are given by simple expressions. The curvature of the clothoid is equal to $t/B$, the angle between the normal vector and the vertical axis is $\pi t^2/2$ and the length of a part from $t_b$ to $t_e$ of the clothoid is equal to $\pi B(t_e - t_b)$. In the athletics track the first part of the clothoid is used, i.e. from $t = 0$ to $t = t^*$, where $t^*$ follows from the smooth connection with the circle part.

## 5.4 The length of a clothoidal athletics track

At this point we can construct a track which has continuous curvature. The inside of the track consists of straight parts, clothoidal parts and circle parts. Now we want to calculate the length of lines that are parallel to the inside of the track. Such lines are called *offset curves*. The offset curves of a straight line are straight lines and the offset curves of a circle part are again circle parts. It can be shown however that the offset curves of a clothoid are not again clothoids. The shape of these curves is given in paragraph 5.4.1. The length of the offset curves of a clothoidal part are calculated in paragraph 5.4.2.

### 5.4.1 The offset curves of the clothoid

Let $\underline{f}_B(t)$ be a clothoid spline with parameter $B$ and let $\underline{n}(t)$ be the normal to this spline

$$\underline{n}(t) = \begin{pmatrix} \sin(\frac{1}{2}\pi t^2) \\ -\cos(\frac{1}{2}\pi t^2) \end{pmatrix}.$$

The offset curve at a distance $d$ of the spline is given by

$$\underline{f}_B(t,d) = \underline{f}_B(t) + d\,\underline{n}(t).$$

This curve is well defined, although it is difficult to see how this curve will look like. It is possible to calculate the length of a part of the offset curve and also the curvature of the offset curve can be found.

### 5.4.2 The length and curvature of the offset curves

A part from $t = 0$ to $t = t_1$ of the offset curve at a distance $d$ of a clothoid with parameter $B$ has length

$$s_B(t_1,d) = \int_0^{t_1} \left\| \frac{\partial \underline{f}_B(\tau,d)}{\partial \tau} \right\| d\tau = \pi t_1 \left( B + \tfrac{1}{2} d t_1 \right).$$

The curvature at point $t_1$ of a clothoid $\underline{f}_B(t)$ is equal to $t_1/B$. The curve radius in this point is then $B/t_1$. The offset curve at a distance $d$ of this clothoid must have a curve radius that is $d$ higher than the curve radius of the clothoid. The curvature at point $t_1$ of the offset curve is then equal to $t_1/(B + dt_1)$. In point $t^*$ the circle part with curve radius of $36.5 + d$ must be connected smoothly to the offset curve at distance $d$ of the clothoid part. The curve radius of the offset curve in the point $t^*$ is equal to $B/t^* + d$. Now we can find that $t^*$ is equal to $B/36.5$ for all $d$.

When parameter $B$ is given, the clothoidal track is uniquely defined. However, before a track can be constructed, two Fresnel integrals must be calculated. This should be done numerically. The calculation of the starting spots on these tracks proceeds in the same way as for the regular tracks. The track is devided in two equal half tracks. The length of one part is calculated and subtracted from the athletic distance $D$ as many times as possible. The remaining number can lie in one of four different intervals that correspond with respectively a straight, a clothoid part, a circle part or another clothoid part. It is not difficult now to calculate the starting spots. However, to mark the spots requires a convenient co-ordinate system. We will now introduce such a system.

## 5.5 How to indicate the starting spots

To indicate the starting spots we define a polar co-ordinate system $(r,\varphi)$ where the polar origin $r = 0$ can be moved over a straight line to a point $x$. The origin $x = 0$, $r = 0$ of this system is chosen in the centre of the track. The $x$-axis is taken parallel to the track straights. The angle $\varphi$ is positive in anti-clockwise direction. $\varphi = 0$ corresponds with the normal to the track at the finish line.

The regular track can very easily be described in this system. To describe the clothoidal track in the new co-ordinate system is not so straightforward. The co-ordinates can only be found by solving numerically the Fresnel integrals that represent the clothoids. The advantage of this co-ordinate system is that lines that are normal to the lanes can be calculated, pointed out and found in a not to complicated way.

# 6 Results

In this chapter, the starting spots on three tracks are calculated. The first two track are regular athletics tracks with curve radii of 36.5 and 40 meter respectively. The last track is an improved track that consisits of clothoidal parts.

**Regular athletics track $R$ = 36.5 meter.**

The curve lengths $s_i$ of the different lanes with width $w$ = 1.25 meter are for this radius equal to

| $s_1$ | $s_2$ | $s_3$ | $s_4$ | $s_5$ | $s_6$ | $s_7$ | $s_8$ |
|---|---|---|---|---|---|---|---|
| 115.61 | 119.54 | 123.46 | 127.39 | 131.32 | 135.25 | 139.17 | 143.10 |

**table 6.1** *Curve length in meters of the lanes of a regular track with radius $R$ = 36.5 meter and lane width $w$ = 1.25 meter.*

The length of the straigths is $L$ = 84.389 meter. The spot for the 200 meter in lane one must be placed exactly at the beginning of the turn before the home straight ($k_1$ = 1, $e_1$ = 0). The starting spots for the 200 meter in lane two up to lane eight lie in this turn ($k_i$ = 0 for $i$ = 2..8). The angles $\varphi$ are equal to

$$\varphi_i = 2\pi - \frac{36.8\,\pi}{36.8 + 1.25(i-1)} = \frac{36.8 + 2.5(i-1)}{36.8 + 1.25(i-1)}\pi.$$

For the 400 meter, the starting spot in lane one coincides with the finish line. The starting spots in the other seven lanes lie in the turn after the home straight ($k_i$ = 1 for $i$ = 2..8). The angles become in this case

$$\varphi_i = \frac{2.5(i-1)}{36.8 + 1.25(i-1)}\pi.$$

The co-ordinates of the starting spots for this track are given in table 4.1.

**Regular athletics track $R = 40$ meter.**

The curve lengths of the different lanes are in this case

| $s_1$ | $s_2$ | $s_3$ | $s_4$ | $s_5$ | $s_6$ | $s_7$ | $s_8$ |
|---|---|---|---|---|---|---|---|
| 126.61 | 130.53 | 134.46 | 138.39 | 142.31 | 146.24 | 150.17 | 154.10 |

**table 6.2** *Curve length in meters of the lanes of a regular track with radius $R = 40$ meter and lane width $w = 1.25$ meter.*

It follows that the length of the straights is equal to $L = 73.394$ meter. The starting spot for the 200 meter in lane one must again be placed at the beginning of the turn before the home straight. The angles of the starting spots in the other lanes are

$$\varphi_i = \frac{40.3 + 2.5(i-1)}{40.3 + 1.25(i-1)}\pi.$$

For the 400 meter the starting spot in lane one again coincides with the finish line. The spots in the other lanes must be placed at the angle

$$\varphi_i = \frac{2.5(i-1)}{40.3 + 1.25(i-1)}\pi.$$

The co-ordinates of the starting spots for this track are given in table 4.2.

**Improved athletics track.**

In this example we take $R = 36.5$ meter, $w = 1.25$ meter and $B = R/2 = 18.25$ meter. In this case the clothoid part starts at parameter $t = 0$ and ends at $t = t^* = 0.5$. The circle part of the turn makes an angle of $3\pi/4$. The lengths of the different parts of the lanes are given in table 6.3. The length of the straights is calculated by subtracting the curve length of lane one from the track length of 400 meters. This length is equal to $L = 55.722$ meter.

| lane | straight part | clothoidal part | circle part |
|---|---|---|---|
| 1 | 55.722 | 28.785 | 86.708 |
| 2 | 55.722 | 29.276 | 89.653 |
| 3 | 55.722 | 29.767 | 92.598 |
| 4 | 55.722 | 30.257 | 95.544 |
| 5 | 55.722 | 30.748 | 98.489 |
| 6 | 55.722 | 31.239 | 101.434 |
| 7 | 55.722 | 31.730 | 104.379 |
| 8 | 55.722 | 32.221 | 107.325 |

**table 6.3** *Lengths in meters of parts of the lanes of a clothoidal athletics track with radius $R = 36.5$ meter and parameter $B = 18.25$.*

The starting spots in lane one for the 200 meter and the 400 meter should be placed again at the end of the back straight and at the end of the home straight respectively. The other starting spots for the 200 meter all lie on the first clothoidal part of the curve before the home straight. The correct angle $\varphi$ can be found by solving for $t_i$ the equation

$$s_{18.25}(t_i,\, d_i) = L + s_i + 2\, s_{18.25}(\tfrac{1}{2},\, d_i) - 200,$$

where

$$d_i = 0.30 + 1.25(i-1),$$

and $s_i$ is the length of the circle part of lane $i$. We are interested in the positive root of the quadratic equation. Values for the angle $\varphi$, the distance $x$ and the radius $r$ are found by substitution of $t_i$ into the following equations

$$\varphi = \left(1 + \tfrac{1}{2}t_i^2\right)\pi\,,$$

$$x = \tfrac{1}{2}L + \pi B \int_0^{t_i} \cos\tfrac{1}{2}\pi u^2\, du - \tan\tfrac{1}{2}\pi t_i^2 \left( R \sin\tfrac{3}{8}\pi + \pi B \int_{t_i}^{\frac{1}{2}} \sin\tfrac{1}{2}\pi u^2\, du \right),$$

$$r = \frac{1}{\cos\tfrac{1}{2}\pi t_i^2} \left( R \sin\tfrac{3}{8}\pi + \pi B \int_{t_i}^{\frac{1}{2}} \sin\tfrac{1}{2}\pi u^2\, du \right).$$

A derivation of these expressions is not given here. The equations follow from geometrical relations.

For the 400 meter, we can find that the starting spots in the lanes two, three and four must be placed on the first clothoid part of the curve after the finish line, whilst the spots for the other four lanes lie on the circle part of the curve after the finish line. The angles $\varphi$, the distance $x$ and the radius $r$ can be found in a way similar to the calculations above.

# 7 Conclusions and recommendations

It is possible to construct an athletics track with continuous curvature which should show improvement compared with regular tracks. However, although the length of these tracks can readily be found, it is difficult to mark the starting spots on these tracks. The lines and marks on an official track have to be renewed at least every three years. Thus the improved athletics track will be very expensive when compared with the regular tracks. Furthermore, the athletes are not convinced in the improvement of the new tracks. The improvement of the track is therefore hypothetical.

The concept of clothoidal tracks might be more useful in indoor tracks. These 200m tracks have sharper turns than the 400m outdoor tracks. The need of smooth curvature is therefore more important for indoor tracks than it is for outdoor tracks.

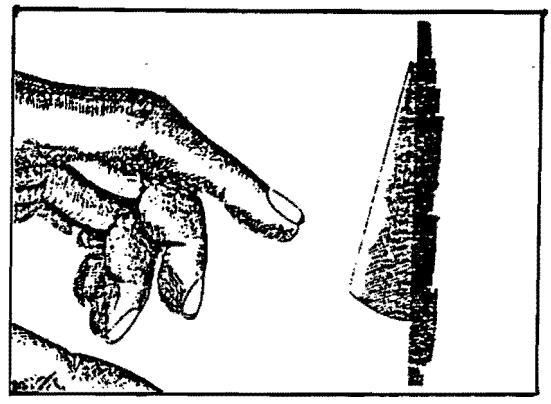
# Reference

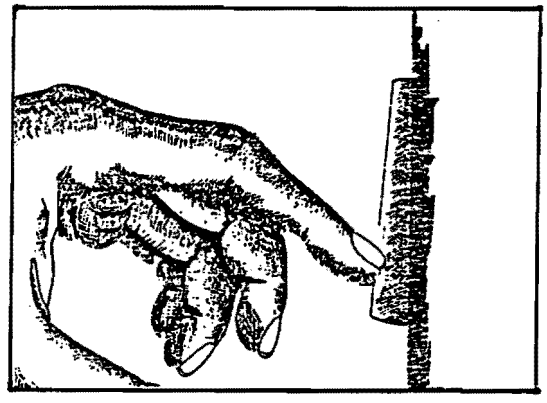
[1] D S Meek and R S D Thomas, *A guided clothoid spline*, Computer Aided Geometric Design 8 (1991) 163-174.

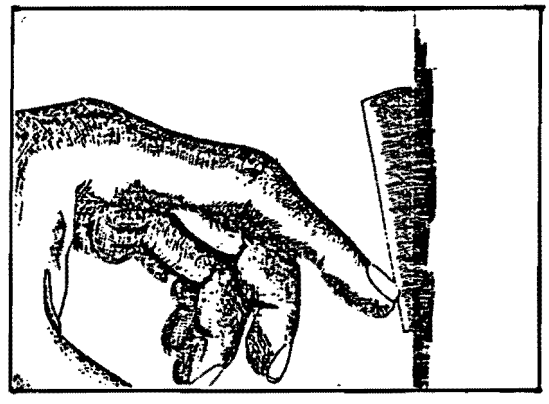
# MICRO-SWITCH OPERATION

**Abstract**

A switch is studied of the type which is often connected to floater/lever combinations. In some situations it rapidly snaps a number of consecutive times from "on" to "off" and vice versa times, while this is not its desired behaviour.
The switch is modelled, and a program is written describing its dynamical behaviour. The input of the program essentially is the force applied onto the switche's plunger; its main output is a function describing the "on-off-behaviour" of the switch as a function of time.
Using this program, the situations in which rapid, unneccesary switchings may occur are spotted. Suggestions are given for improving the design of the switch; they are tested by means of the same program.
It actually seems to be possible to redesign the switch in such a way that the aforementioned switchings are avoided.







klik...

# Micro-Switch Operation

## Introduction

You are driving along the highway in your car. It is past midnight, and all gas stations are closed – but who cares: the indicator on your dashboard shows that you have plenty of fuel. At least, it did a few seconds ago ! Suddenly, the indicator has dropped down to the red area, indicating that you are running out of gas ! You seem to have a serious problem, but luckily the indicator moves upward after a few kilometers – and unluckily, it moves down again two minutes later... Are you starting to feel insecure, and do you wonder what is going on ? Read this report !

The fuel indicator on the dashboard is controlled by a micro-switch, which is connected to a floater/lever combination. The floater floats on the fuel in the tank. If the fuel level decreases, the force the floater/lever combination exerts on the switch increases, and at a certain moment it is activated, resulting in a downward movement of the indicator.
A certain force must be applied to the micro-switch to activate it. Assume that the fuel level in the tank is such that the floater/lever combination applies exactly that force to the plunger of the switch. Theoretically, the only thing that happens is that the switch is activated. However, when the car makes a turn or when the road is bad, the fuel in the tank moves, and so does the floater. As a consequence, the fuel level and the floater/lever force vary; the switch rapidly move from the "on"-position to the "off"-position and vice versa, which explains the peculiar movements of the indicator on your dashboard.

It is clear that such rapid and unnecessary oscillations should be avoided. In this report, a particular micro-switch is considered. It is modeled in order to find out what modifications could be made to improve it, and to make sure that you feel secure in your car even at times when the gas stations are closed...

The outline of this report is as follows.
In its first part, we will describe the switch and we will roughly consider possible causes of rapid switchings. This will yield a rather general problem formulation.
The second part deals with modelling the switch, and with the translation of the general problem into a more precise and mathematically formulated one.
A solution method for this mathematical problem will be discussed in the third part of this report. We will get to an implementation in Turbo-Pascal: given the forces applied on the plunger, we can predict the behaviour of the (modeled) switch.
The last topic we will consider is the improvement of the micro-switch. We will not try to devise a new kind of switch; we will restrict ourselves to indicating which parameters of the switch – e.g. plunger length, spring constant – may be altered to avoid unnecesary rapid oscillations.

# 1 The Switch and Its Problem

## 1.1 The Switch

The micro-switch we will consider is shown in Figure 1. Its essential parts are the plunger, the carrier and the spring, the contact levels, and the lever: their (relative) positions determine whether the switch is on or off.

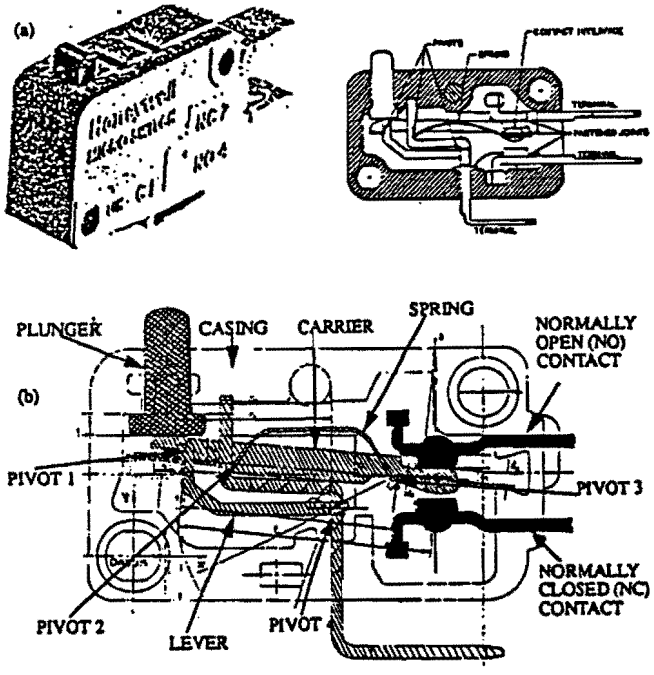



**Figure 1:** The micro-switch

A switching sequence of the micro-switch is displayed in Figure 2. It is the series of events that occur in a fuel-controlling switch when the fuel level decreases. Just after tanking fuel, no force is actuated on the plunger (a). When the fuel level decreases, the actuating force increases and moves the carrier of the switch to the equilibrium position: the switch is off (b), but even if the force augments infinitely little, it will snap to the on-position (c). If the force increases even more, the switch stays on (d).

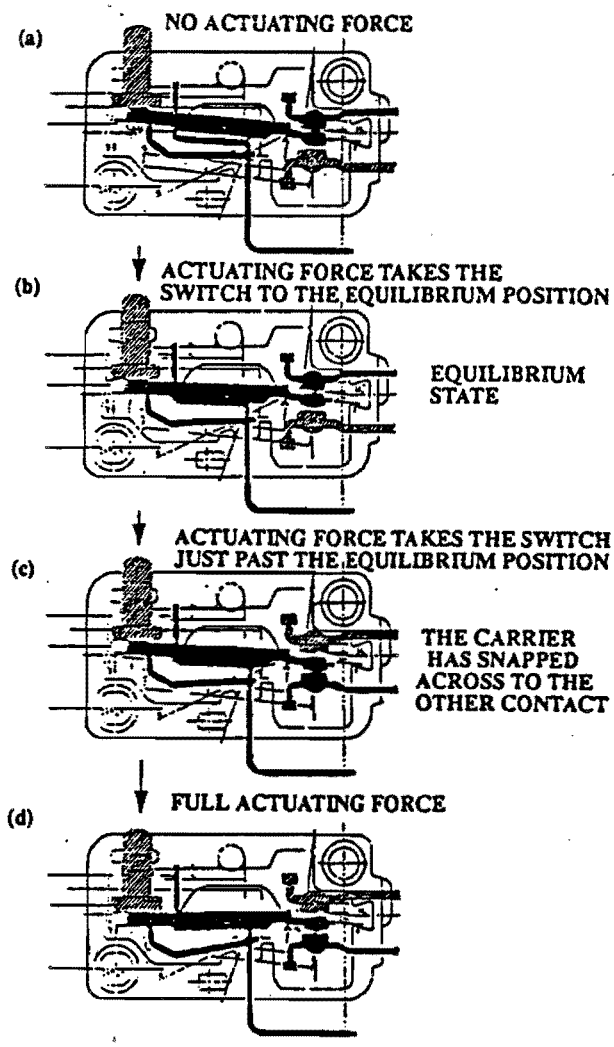



**Figure 2:** A switching sequence

## 1.2 The Switch's Problem

Even if the fuel level and the corresponding force the floater/lever combination applies on the switch plunger take the carrier (close) to its equilibrium position, it must snap from off to on at once, without any unnecessary switchings. Intuitively speaking, we want to devise a fuel indicator which is not influenced by curves in and bumps on the road.
In Figure 2, the rapid switchings we deal with in this report occur between states (b) and (c).

# 2 The Modelling

## 2.1 Essentials of the Switch

As mentioned earlier, the main parts of the micro-switch are the plunger, the carrier and the spring, the contact levels, and the lever. If we forget about the case the switch is in, and about the actual shape of the different components, the micro-switch can be represented schematically as in Figure 3.

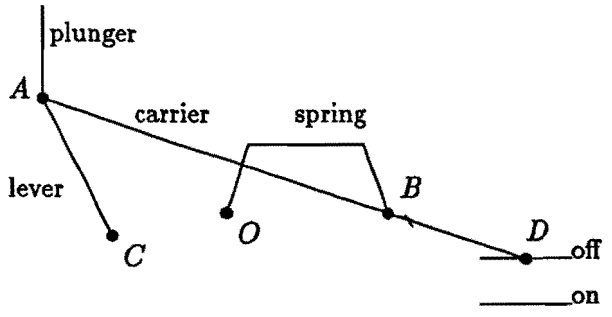



**Figure 3:** Schematical representation of the micro-switch

A number of implicit assumptions have been made to come to this switch representation; some others seem useful to avoid unnecessary complication of the problem.

Assumptions about movements:

- The left side of the spring is fixed at point $O$. It is allowed to rotate freely around this point.
- The right side of the spring is fixed at point $B$, which is a fixed point on the carrier.
- The carrier is connected to the lever in point $A$. Rotation of the carrier around the lever in $A$ is allowed.
- The right end of the carrier $D$ can move in horizontal direction along the off- or on-level.
- The lever is fixed at point $C$. It is allowed to rotate freely around this point.

- The carrier can move "through" point $O$, and "through" the spring. Although this may seem strange from Figure 3, the construction of the switch displayed in Figure 1 explains this assumption. Similarly, the carrier can move "through" the lever.

Assumptions about construction:

- No frictions occur between any two parts of the switch.
- There is no play in the sizes of the different switch parts (except, of course, in the spring).
- The plunger contacts the carrier in exactly one point.
- The off- and on-levels are fixed. They correspond to the off- and the on-position of the switch respectively.

Assumptions about forces:

- We neglect gravitational forces.
- When no force is applied on the plunger, the switch is in its off-position. In that situation, the spring is already stressed, i.e. the length $d$ of $OB$ is smaller than in the equilibrium situation of the spring.
- In all positions, the spring behaves as an ordinary spring with middle line along $OB$: the force the spring applies is directed along $OB$.
- In $A$, a vertical force can be applied by pushing the plunger downward.

When does the spring switch from off to on ?
According to the last assumption about forces, the spring makes the carrier move in the direction of $OB$ at all times. If, in Figure 3, $O$ lies beneath the carrier, the spring applies a force which is directed upward, and the switch is in its off-position. If $O$ lies above the carrier, the spring pushes the carrier down, and the switch is in its on-position. Only if $O$ lies on $AD$, the spring force is directed along the carrier, which may therefore move upward or downward.
This reasoning justifies to the following assumption:

Assumption about forces:

- The rightmost end point $D$ of the carrier will move from the off-level to the on-level or vice versa exactly when the the spring force is directed along the carrier, i.e. when $O$ is on $AD$ in Figure 3.

## 2.2 Restatement of the Problem

Under the mentioned assumptions about movements within the switch, construction of the switch, and forces, we can reformulate the original problem more mathematically.

We wish to find out how the switch behaves if a given force is applied on the plunger. We are interested in both the statical and the dynamical behaviour.
As far as the statical behaviour of the switch is concerned, we want to know what force has to be applied on the plunger to keep the system in equilibrium, given the vertical displacement of A. If we know this, it seems to be quite easy to derive the dynamical behaviour as a rapid consecution of statical situations.

Finding a description of the dynamical behaviour of the switch is the mathematical problem we will be concerned with. The practical problem will be to use this behaviour description for improvement of the switch: even if the force applied on the plunger takes the carrier (close) to its equilibrium position, it must snap from off to on at once, without unnecessary switchings.

## Parameters

| | |
|---|---|
| $d$ | distance between $O$ and $B$ (m); |
| $d_0$ | distance between the end points of the spring in the equilibrium position (m); |
| $k$ | spring constant (N/m); |
| $s$ | distance between $A$ and $B$ (m); |
| $R$ | distance between $A$ and $C$ (m); |
| $L$ | length of the carrier (m); |
| $x_i$ | $x$-coordinate of $i$ according to the origin $O$ ($i \in \{A, B, C, D\}$); |
| $y_i$ | $y$-coordinate of $i$ according to the origin $O$ ($i \in \{A, B, C, D, \text{off}, \text{on}\}$). |

# 3 Predicting the Switche's Behaviour

We will consecutively derive mathematical descriptions of the statical and the dynamical behaviour of the switch. This will yield a program in which the force applied onto the plunger as a function of time can be input, and the output of which comprises whether the switch is on or off at a certain moment in time. We will use this program to find out when exactly the rapid, unneccesary oscillations occur.

## 3.1 Analysis of the Statical Situation

In this section we will focus on finding the force that we have to apply on the plunger to keep the carrier in a statical equilibrium position. We assume to know the actual position of the carrier, the positions of $A$ and $B$, and whether the switch is on or off.

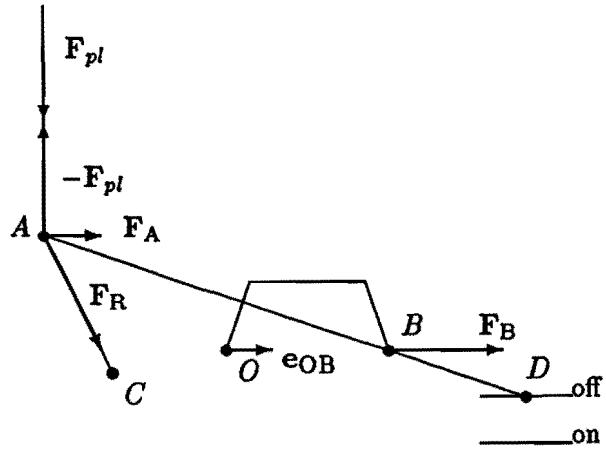



**Figure 4:** Determining the force to be applied on the plunger to maintain equilibrium: the main idea

Suppose the spring applies a force $\mathbf{F}_B$ on the carrier in $B$. This force is directed along $OB$:

$$\mathbf{F}_B = k.(d - d_0).\mathbf{e}_{OB},$$

where $\mathbf{e}_{OB}$ is the two dimensional unit vector directed from $O$ to $B$. We want to know what force we have to apply onto the plunger of the switch to keep the system in equilibrium.

Let the switch be in its off-position, i.e. $y_D = y_{off}$. The force $\mathbf{F}_B$ working in $B$ now causes a force $\mathbf{F}_A$ in $A$. $\mathbf{F}_A$ can be determined from $\mathbf{F}_B$. We will do so explicitly in the sequel.
From $\mathbf{F}_A$ we can derive the force that has to be applied on the plunger to keep the system in equilibrium. It suffices to decompose $\mathbf{F}_A$ into a force $\mathbf{F}_R$ directed along $CA$, and a force $-\mathbf{F}_{pl}$ directed along the plunger. The force that has to be applied to maintain equilibrium then equals $\mathbf{F}_{pl}$; The force $\mathbf{F}_R$ will be compensated by a reaction force, because of the connection between $A$ and $C$ by the lever.
$\mathbf{F}_{pl}$ can be determined in a similar way if the switch is in its on-position; only some details in the calculations will change.

We will now determine the aforementioned forces explicitely in the case of Figure 5. There are three other cases which can be handled in a similar way; we refer to the computer program for the statical situation which can be found in Appendix A.

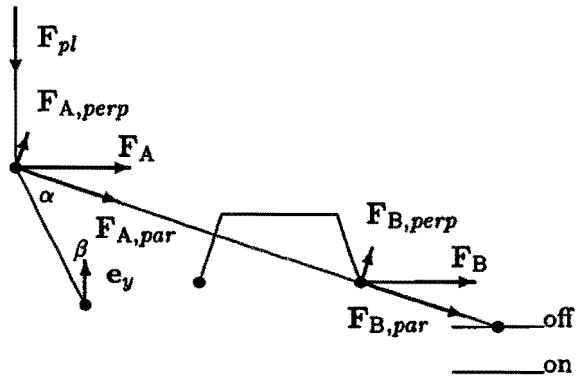



**Figure 5**: Determining the force to be applied on the plunger to maintain equilibrium: explicit calculations

Let $\mathbf{r}_A$ and $\mathbf{r}_B$ be the two-dimensional position vectors of $A$ and $B$. As $AC$ has length $R$, we have

$$\mathbf{r}_A = (x_A, y_A), \text{ where} \tag{1}$$

$$x_A = x_C - \sqrt{R^2 - (y_A - y_C)^2}. \tag{2}$$

Notice that $y_A$ equals the vertical displacement of the plunger !
Similarly,

$$\mathbf{r}_B = (x_B, y_B), \text{ and} \tag{3}$$

$$x_B = x_A + s.\sqrt{1 - \frac{1}{L^2}.(y_{off} - y_A)^2}, \tag{4}$$

$$y_B = y_A + \frac{s}{L}.(y_{off} - y_A). \tag{5}$$

Decomposition of $\mathbf{F}_B$ into a component along the carrier and one perpendicular to it, as in Figure 5, yields

$$\mathbf{F}_{B,par} = \left(\mathbf{F}_B, \frac{\mathbf{r}_B - \mathbf{r}_A}{|\mathbf{r}_B - \mathbf{r}_A|}\right).\frac{\mathbf{r}_B - \mathbf{r}_A}{|\mathbf{r}_B - \mathbf{r}_A|}, \tag{6}$$

$$\mathbf{F}_{B,perp} = \left|\mathbf{F}_B \times \frac{\mathbf{r}_B - \mathbf{r}_A}{|\mathbf{r}_B - \mathbf{r}_A|}\right|.\mathbf{e}_{perp}. \tag{7}$$

Here, $(\mathbf{u}, \mathbf{v})$ denotes the inner product of two two-dimensional vectors $\mathbf{u}$ and $\mathbf{v}$, and $\mathbf{u} \times \mathbf{v}$ denotes their vector product. $|\mathbf{u}|$ equals the 2-norm of $\mathbf{u}$. The unit vector $\mathbf{e}_{perp}$ is the unit vector perpendicular to the carrier in the direction shown in Figure 5. The force $\mathbf{F}_A$ applied in $A$ caused by $\mathbf{F}_B$ can be decomposed in a similar way into a component along and a component perpendicular to the carrier:

$$\mathbf{F}_A = \mathbf{F}_{A,par} + \mathbf{F}_{A,perp}, \text{ where} \tag{8}$$

$$\mathbf{F}_{A,par} = \mathbf{F}_{B,par} \tag{9}$$

$$\mathbf{F}_{A,perp} = \frac{s}{L}.\mathbf{F}_{B,perp}. \tag{10}$$

If $\mathbf{r}_{AC}$ is the vector with length from $A$ to $C$, we have from Figure 5:

$$\alpha = \arccos\left(\left(\frac{\mathbf{r}_{AC}}{|\mathbf{r}_{AC}|}, \frac{\mathbf{F}_A}{|\mathbf{F}_A|}\right)\right), \text{ and} \tag{11}$$

$$\beta = \arccos\left(\left(\frac{-\mathbf{r}_{AC}}{|\mathbf{r}_{AC}|}, \mathbf{e}_y\right)\right). \tag{12}$$

Here, $\mathbf{e}_y$ is the unit vector in the positive $y$-direction.

Using the sinus rule, we eventually find an expression for the force to be applied on the plunger to maintain the equilibrium position of the switch:

$$\mathbf{F}_{pl} = -\frac{|\mathbf{F}_A|.\sin(\alpha)}{\sin(\beta)}.\mathbf{e}_y \tag{13}$$

Notice that this expression is only valid if the switch is in its off-position, i.e. if $y_D = y_{off}$. If the switch is on $y_{off}$ has to be replaced with $y_{on}$ in the above formulae.

It is necessary to find a mathematical criterion to determine whether the switch is in its off- or in its on-position. We assumed that the switch moves from off to on when the origin $O$ lies on $AB$, i.e. when

$$y_A + \frac{|\mathbf{r}_A|}{L}.(y_{off} - y_A) = 0 \tag{14}$$

It switches from on to off again when

$$y_A + \frac{|\mathbf{r}_A|}{L}.(y_{on} - y_A) = 0 \tag{15}$$

Thus we have found expressions for the force to be applied on the plunger of the switch to keep the system in its equilibrium position, given that the vertical displacement of the plunger equals $y_A$ and given the on- or off-position of the switch. A graph showing the relation between $y_A$ and $\mathbf{F}_{pl}(y_A)$ is depicted in Figure 6.
We have chosen

| | | | |
|---|---|---|---|
| $d_0$ | 0.012 m; | $y_{A,0}$ | 0.004 m; |
| $k$ | 150 N/m; | $x_C$ | 0.005 m; |
| $s$ | 0.015 m; | $y_C$ | -0.002 m; |
| $R$ | 0.010 m; | $y_{on}$ | -0.001 m; |
| $L$ | 0.020 m. | $y_{off}$ | 0.001 m; |

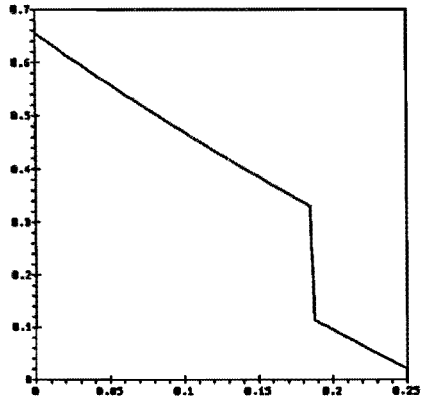


**Figure 6:** The force to be applied to the plunger as a function of the displacement $y_A - y_{A,0}$; the switch is off initially. The force is displayed in Newton along the vertical axis; the displacement in centimeters along the horizontal one

This figure is the output of a computer program in Turbo-Pascal that can be found in Appendix A. The main part of the program is a procedure that computes $\mathbf{F}_{pl}(y_A)$ from $y_A$, if it knows whether the switch is off or on. The program body calls this procedure for values of $y_A$ within a given range. This same procedure determines also the new state of the switch (on or off) if the plunger displacement equals $y_A$, because it could be that the carrier has gone through the origin $O$.
We will not go into details as far as the program is concerned; the code is abundantly annotated.

Notice, however, that the graph displayed in Figure 6 has been obtained assuming the switch to be in its off-position initially. We see that the consists of three parts; the steepest part reflects the switching from off to on. Slightly different results appear when the switch is on in the beginning; they are shown in Figure 7. The figure looks alike the previous one; only the steepest part has moved a little bit. This means that the switching from on to off happens for an other vertical displacement.

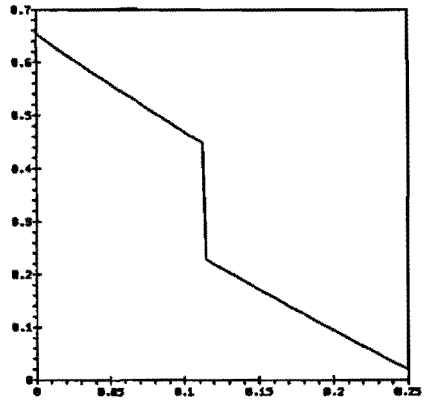


**Figure 7**: The force to be applied to the plunger as a function of the displacement $y_A - y_{A,0}$; the switch is on initially. The force is displayed in Newton along the vertical axis; the displacement in centimeters along the horizontal one

## 3.2 Analysis of the Non-Statical Situation as Rapid Consecution of Statical Situations

What happens with the switch if a dynamical force is applied to the plunger ?

Let the dynamical force be $\mathbf{F}(t)$, and let $y_A(t)$ be the displacement of the plunger at time $t$. We then have

$$M.\ddot{y}_A(t)\mathbf{e}_y = \mathbf{F}(t) - \mathbf{F}_{pl}(y_A(t)) \text{ or} \tag{16}$$

$$y_A(t)\mathbf{e}_y = \int_0^t \int_0^x \frac{\mathbf{F}(z) - \mathbf{F}_{pl}(y_A(z))}{M} dz dx, \tag{17}$$

where $M$ denotes the mass of the plunger.

It is not difficult to approximate $y_A(t)$ numerically by discretizising the above integral equation.

Assume we apply the force $\mathbf{F}(t)$ for $t \in [0,T]$. We will divide this time interval into $N$ small time steps: we will approximate $y_A(t)$ at time points $t_n = n.T/N$, $n \in \{0,1,\ldots,N\}$.
If the velocity $v_n$ of the plunger at time point $t_n$ is known, its position at $t_{n+1}$ can be approximated. Denote by $y_{A,n}$ the $y$-coordinate of $A$ at time $t_n$, then

$$y_{A,n+1} = y_{A,n} + v_n.\delta, \tag{18}$$

where $\delta = \frac{T}{N}$. The acceleration $a_n$ equals $\ddot{y}_A$ at time $t_n$, so

$$\begin{aligned} a_n \mathbf{e}_y &= \ddot{y}_A(t_n)\mathbf{e}_y && (19)\\ &= M.(\mathbf{F}(t_n) - \mathbf{F}_{pl}(y_A(t_n))) && (20)\\ &= M.(\mathbf{F}(t_n) - \mathbf{F}_{pl}(y_{A,n})) && (21) \end{aligned}$$

So if we know $y_A$ and the position of the switch we can compute $a_n$. Hence we can approximate $v_n$ by

$$v_n = \frac{1}{2}.a_0 + a_1 + \cdots + a_{n-1} + \frac{1}{2}.a_n \tag{22}$$

As we know $\mathbf{F}(t)$, we also know $\mathbf{F}$ at the discrete time points $t_n$. Furthermore, we know the initial value of $y_A$, i.e. $y_{A,0}$. From the above formulae, we can successively determine $a_0$, $y_{A,1}$, $a_1$, $v_1$, $y_{A,2}$, $a_2$, $v_2$, $y_{A,3}$ etc.. The values of $\mathbf{F}_{pl}(y_{A,n})$ can be computed using the results of the previous section. Notice that at each time point $t_n$ we have to know whether the switch is on or off. We have already described a mathematical criterion to find this out, and programmed in the code presented in Appendix A. The same method may be used here.

A second computer program in Turbo-Pascal has been written to determine $y_A(t)$, given the dynamical force $\mathbf{F}(t)$ that is applied on the plunger of the switch. The program uses the aforementioned discretization. Furthermore, at each time point it checks whether the switch is on or off. As a side result it therefore yields the "on-off-behaviour" of the switch, given the force applied on the plunger.
The program code is included in Appendix B.

### 3.3 Output of the Behaviour Predicting Program: When Exactly do the Unneccesary Rapid Switchings Occur ?

In this section we present some output results of the switch behaviour predicting program. We will concentrate on studying the relation between the force applied on the plunger of the switch, and its "on-off-behaviour". In fact, as our principal goal is to avoid unneccesary rapid oscillations – like the ones occurring in the fuel indicator –, this "on-off-behaviour" is of particular interest.

Figure 8a and 8b show a constant force applied on the plunger of the switch, and its corresponding "on-off-behaviour". As one would expect, the switch does not switch in this case. In the fuel indicator example, this situation corresponds with a car that does not move, the engine being turned off. In that case the fuel indicator is not quite likely to move.

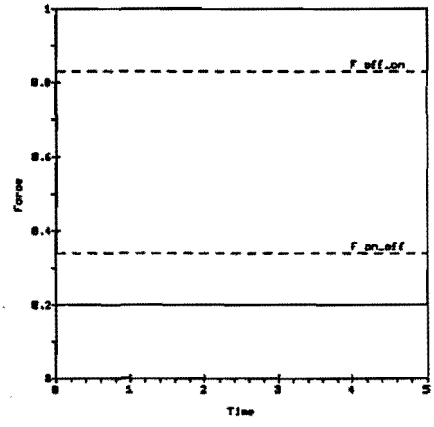



**Figure 8a:** A constant force (in Newton) applied on the plunger of the switch; time is indicated in seconds

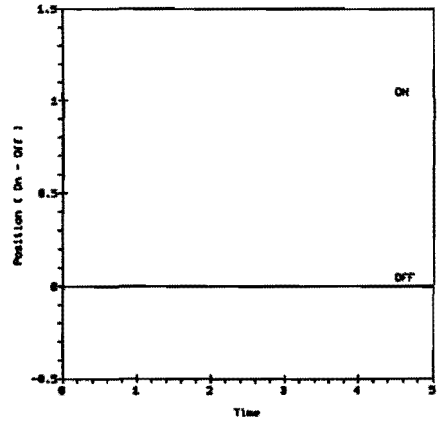



**Figure 8b:** "On-off-behaviour" of the switch if a constant force is applied onto the plunger; time is indicated in seconds

The situation depicted in Figures 9a and 9b is slightly more realistic. The force applied onto the plunger increases linearly with time, as would happen if the car

does not move, but if its engine is turned on for a long time. The fuel level decreases regularly, and the force applied onto the floater/lever combination increases in a corresponding way.
Figure 9b displays that in this situation the switch snaps from "off" to "on" at once, without unneccesary switchings.

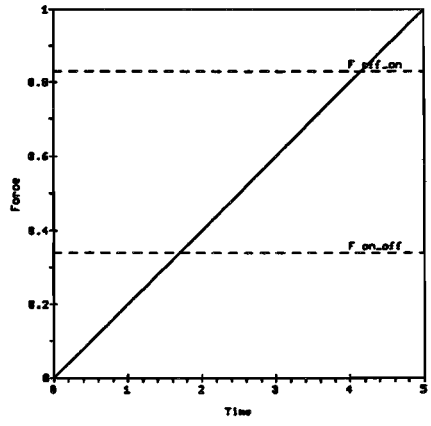



**Figure 9a:** A linear, increasing force (in Newton) applied on the plunger of the switch; time is indicated in seconds

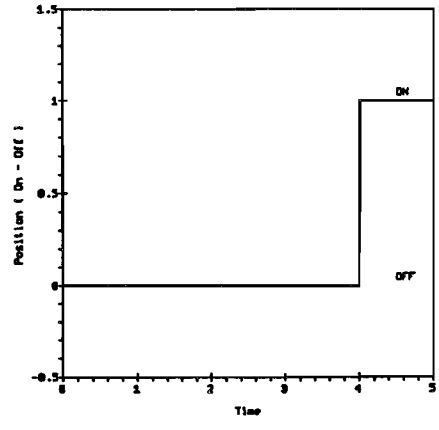



**Figure 9b:** "On-off-behaviour" of the switch if a linear, increasing force is applied onto the plunger; time is indicated in seconds

In real life, people drive in cars. Not only does the fuel level in the tank decrease, but in the same time the car moves. Although one could assume the amount of fuel

to decrease linearly in time, the fuel level – or rather the fuel surface – does not. If the road is bad, the surface is bound to be wavy. The floater will therefore not move down regularly when the fuel amount decreases, but it will move up and down consecutively, the downward movements slightly prevailing.
The force applied onto the switche's plunger will therefore in practice look like in Figure 10a. The corresponding "on-off-behaviour" is displayed in Figure 10b.

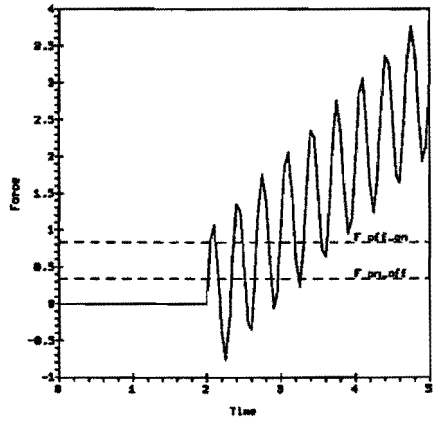



**Figure 10a:** A "wavy" linear, increasing force (in Newton) applied on the plunger of the switch; time is indicated in seconds

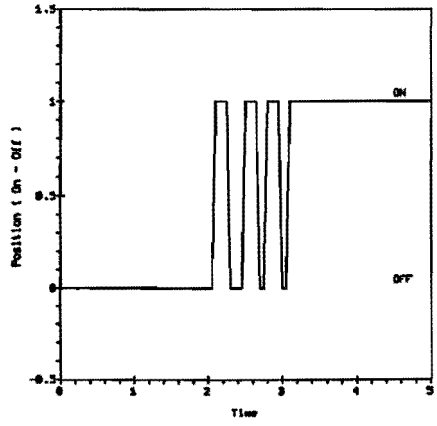



**Figure 10b:** "On-off-behaviour" of the switch if a "wavy" linear, increasing force is applied onto the plunger; time is indicated in seconds

Of course, the situation depicted in the latter pictures is still a simplification of

reality. Nevertheless, Figures 10a and 10b allow us to spot the switche's problem.

If the switch is connected to a floater/lever combination – as for the fuel indicator of a car –, rapid, unneccesary switchings occur if a "wavy" force is applied onto the plunger and if the average size of this force is close to the one needed to let the switch snap from "off" to "on" or vice versa.

One might argue that there is another case in which rapid, unneccesary oscillations occur.
Suppose it takes a time $\delta$ for the switch to actually snap from "on" to "off" or vice versa. If a (sinusoidal) force is applied onto the plunger with a period close to $\delta$, and with an average size close to the force needed to cause a switching, rapid oscillations are bound to occur.
However, in practical situations, $\delta$ is quite small. Taking parameters as have been used in the aforementioned computer programs, $\delta$ is approximately equal to 0.02 seconds. We will assume the force applied onto the plunger to have a period exceeding $\delta$ – which seems to be a quite realistic assumption.

# 4 Solving the Switche's Problem

In this section we will consider the actual problem addressed in this report: how to avoid the switch from rapidly snapping from "on" to "off" and vice versa if there is no real reason to do so.
Using the dynamical switch behaviour predicting program, we found out when exactly these rapid, unneccesary switchings occur. They seem to be mainly caused by a "wavy" plunger force which is close to the force needed to let the switch snap from "on" to "off" or vice versa.

In the sequel we will use the dynamical switch behaviour predicting program to find out what parameters may be changed to overcome the switche's problem.

Two parameters can intuitively be seen to influence the switche's behaviour: the spring constant $k$, and the allowed range of plunger displacement. A larger spring constant makes it harder to snap the switch, so that "waves" in plunger forces have less influence on its behaviour. Making the allowed range of plunger displacement larger increases the difference between the forces needed to let the switch snap from "off" to "on" and vice versa. This also results in less influence of "waves" in forces on the switche's behaviour.

## 4.1 Changing the Spring Constant

If you are driving along the highway around midnight – all gas stations are closed –, and if you do not want to bother with fuel indicators successively indicating you are out of fuel and there is plenty of fuel left in you tank, here is what you could do. Find the switch connected to the floater/lever combination in the tank, and replace its spring with one with infinite spring constant $k$. Practically, this means that the switch will not react any more, whatever force the floater/lever combination applies onto its plunger.

Of course, this is quite a rude way of solving the fuel indicator problem – but it works...

We studied the behaviour of the switch if its spring is replaced by one with a larger spring constant. The results are depicted in Figure 11a and 11b. Notice that the input forces displayed in Figures 10a and 11a are the same. Figure 11b shows that after replacing the spring the rapid switchings no longer occur !

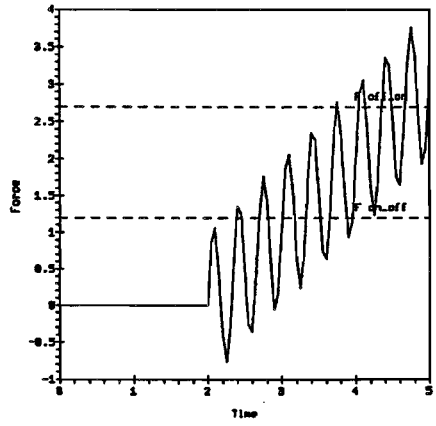



**Figure 11a:** A "wavy" linear, increasing force (in Newton) applied on the plunger of the switch; time is indicated in seconds

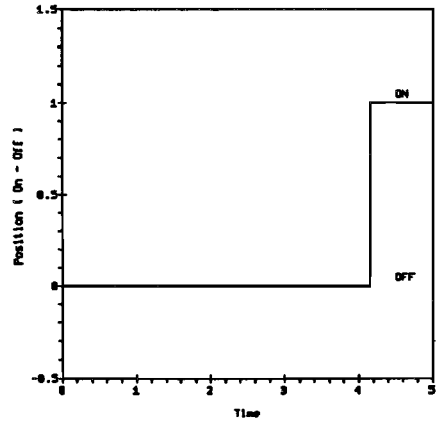



**Figure 11b:** "On-off-behaviour" of the switch if a "wavy" linear, increasing force is applied onto the plunger; time is indicated in seconds. The spring constant has been increased compared to Figure 10b

## 4.2 Changing the Allowed Range of Plunger Displacement

Once you have started "rebuilding" the switch controlling the fuel indicator of your car, you might as well do the job completely.

We used the dynamical switch behaviour predicting program to find out how increasing the allowed range of plunger displacement influences the occurrence of rapid switchings. The results are depicted in Figures 12a and 12b. Figure 12a again shows the "wavy" plunger force introduced in Figure 10a.
Notice that no unneccesary rapid switchings occur after increasing the allowed range of plunger displacement !

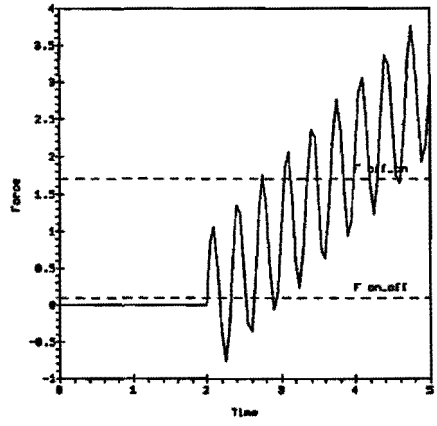



**Figure 12a:** A "wavy" linear, increasing force (in Newton) applied on the plunger of the switch; time is indicated in seconds

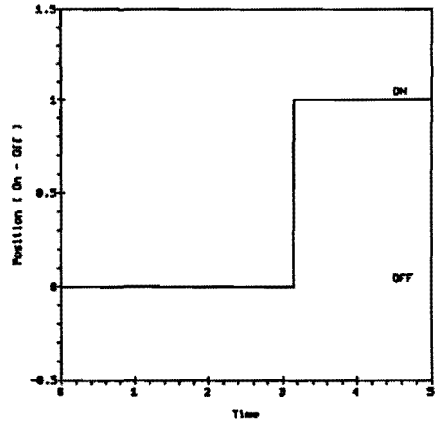



**Figure 12b:** "On-off-behaviour" of the switch if a "wavy" linear, increasing force is applied onto the plunger; time is indicated in seconds. The allowed range of plunger displacement has been increased compared to Figure 10b

# 5 Conclusions

In this report we considered a switch of the type which is used in combination with floater/lever combinations. The lever is connected to the plunger of the switch; if the force applied onto the plunger is large enough, the switch snaps from "off" to "on" or vice versa.
In some situations, however, many consecutive rapid switchings appeared to occur. The practical problem we were asked to solve was to redisign the switch in such a way that these unneccesary switchings were avoided.

In order to grasp this problem, we made a number of assumptions on how the considered switch works and looks like. This yielded a more mathematical problem, the solution of which could be used to solve the practical one: give mathematical descriptions for the statical and dynamical behaviour of the switch. In the statical case, the force applied onto the plunger is assumed to be constant, while in the dynamical case, it is allowed to vary in time.

Using quite elementary fysics, we were able to approximate both the statical and the dynamical behaviour of the switch – under the aforementioned assumptions. For both the statical and the dynamical case, we have written programs in Turbo-Pascal, the first being used as a procedure in the second.
The input of the dynamical behaviour predicting program is essentially the force applied onto the plunger, and the initial position of the switch; the main ouput is a function describing when the switch is "on" and when it is "off" as a function of time.

Using the dynamical switch behaviour predicting program, we found out that rapid, unneccesary switchings mainly occur when a "wavy" force is applied onto the plunger, and if the average size of this force is close to the one needed to let the switch snap from "off" to "on" or vice versa.

Some common sense and verification using the same behaviour predicting program showed which switch parameters had to be changed in order to solve the switche's problem. Increasing the spring constant of the spring contained in the switch, or making the allowed range of plunger displacement larger both seem to be acceptable solutions to avoid unneccesary, rapid switchings.

# Appendix A

```
{ This program computes the force that one has to apply on the
  plunger to keep the system in equilibrium, given the position
  of A                                                          }

{ ------- PROGRAM ------------------------------------------------ }

 PROGRAM plungerforce
{$N+}

{ Filling in the constants found by measurements on a real switch: }

;CONST xc=0.005
;      yc=-0.002
;      l=0.018
;      R=0.01
;      yoff=0.001
;      yon=-0.001
;      S=0.014
;      d0=0.012
;      k=150

{ All the constants are in standard units; meters and Newton/meter }

;TYPE real = extended

;VAR Fpl,ya: real
;    N,on  : integer
;    f     : text


{ ------- PROCEDURES AND FUNCTIONS ------------------------------ }

;FUNCTION ARCCOS(x: real): real
  ;BEGIN IF x<0 THEN ARCCOS:=ArcTan(Sqrt(1-x*x)/x)+pi
  ;      IF x=0 THEN ARCCOS:=pi/2
  ;      IF x>0 THEN ARCCOS:=ArcTan(Sqrt(1-x*x)/x)
  ;END

{ ------------------------------------------------------------------ }

;PROCEDURE PLF(ya: real; VAR on: integer; VAR Fpl: real)

{ This is the main procedure of this program; it computes not only
  the plungerforce, but it checks also whether the position of the
```

```
switch has to be changed and if so it changes the value that
indicates the position of the switch.                              }

;VAR xa,xb,yb,y,d                    : real
;    FBx,FBy,FBa,FBp,ex,ey           : real
;    FAa,FAp,FAx,FAy,alpha,beta,h1,h2: real

;BEGIN

   { the y-coordinate of A should be chosen in some interval }
   { we now compute the x-coordinate of A }
   ;xa:=xc-Sqrt(R*R-(ya-yc)*(ya-yc))

   { testing wether the origin is beneath the carrier }
   ;IF on=1 THEN y:=yon
            ELSE y:=yoff
   ;IF ya+Sqrt(xa*xa+ya*ya)*(y-ya)/l>0 THEN BEGIN y:=yoff;on:=0;END
                                       ELSE BEGIN y:=yon;on:=1;END

   { computation of the coordinates of B }
   ;xb:=xa+S*Sqrt(1-(y-ya)*(y-ya)/(l*l))
   ;yb:=ya+S*(y-ya)/l

   { computation of the force in B }
   ;d:=xb*xb+yb*yb
   ;FBx:=k*(1-d/d0)*xb
   ;FBy:=k*(1-d/d0)*yb

   { its components perpendicular to and along the carrier are
     given by:                                                     }
   ;ex:=(xb-xa)/S
   ;ey:=(yb-ya)/S
   ;FBa:=FBx*ex+FBy*ey
   ;FBp:=ex*FBy-ey*FBx

   { the components of the force in A perpendicular to and along
     the carrier are given by:                                     }
   ;FAa:=FBa
   ;FAp:=S*FBp/l

   { and its (x,y)-coordinates are: }
   ;FAx:=FAa*ex-FAp*ey
   ;FAy:=FAa*ey+FAp*ex

   { we now compute the magnitude of the plungerforce which
     with the reaction force compensates the force at A.
     We have to make distinction between four different cases      }
   ;h1:=xc-xa
   ;h2:=yc-ya
   ;IF h1*FAy-h2*FAx>=0
    THEN IF FAx>=0
```

```
          THEN BEGIN alpha:=ARCCOS((h1*FAx+h2*FAy)/
                                   (Sqrt(h1*h1+h2*h2)
                                  *Sqrt(FAx*FAx+FAy*FAy)))
               ;     beta:=ARCCOS(-h2/Sqrt(h1*h1+h2*h2))
               ;     Fpl:=Sin(alpha)
                              *Sqrt(FAx*FAx+FAy*FAy)/
                               Sin(beta)
               ;END
          ELSE BEGIN alpha:=ARCCOS((-h1*FAx-h2*FAy)/
                                   (Sqrt(h1*h1+h2*h2)
                                  *Sqrt(FAx*FAx+FAy*FAy)))
               ;     beta:=ARCCOS(h2/Sqrt(h1*h1+h2*h2))
               ;     Fpl:=Sin(alpha)
                              *Sqrt(FAx*FAx+FAy*FAy)/
                               Sin(beta)
               ;END
     ELSE IF FAx>=0
          THEN BEGIN alpha:=ARCCOS((h1*FAx+h2*FAy)/
                                   (Sqrt(h1*h1+h2*h2)
                                  *Sqrt(FAx*FAx+FAy*FAy)))
               ;     beta:=ARCCOS(h2/Sqrt(h1*h1+h2*h2))
               ;     Fpl:=-Sin(alpha)
                              *Sqrt(FAx*FAx+FAy*FAy)/
                               Sin(beta)
               ;END
          ELSE BEGIN alpha:=ARCCOS((-h1*FAx-h2*FAy)/
                                   (Sqrt(h1*h1+h2*h2)
                                  *Sqrt(FAx*FAx+FAy*FAy)))
               ;     beta:=ARCCOS(-h2/Sqrt(h1*h1+h2*h2))
               ;     Fpl:=-Sin(alpha)
                              *Sqrt(FAx*FAx+FAy*FAy)/
                               Sin(beta)
               ;END
 END

{ ------- MAIN PROGRAM ---------------------------------------- }

;BEGIN
 assign(f,'dat2.dat')
;rewrite(f)
;on:=0
;FOR N:=0 TO 100 DO BEGIN ya:=0.0015-N*0.00003
                    ;     PLF(ya,on,Fpl)
                    ;     writeln(f,100*(0.0015-ya),'  ',Fpl)
                    END
;close(f)
END.
```

# Appendix B

```
{ This program simulates the behaviour of the switch given the
  force in time that one applies on the plunger.                    }

{ ------- PROGRAM ------------------------------------------------ }

PROGRAM plungerforce
{$N+}

{ The working of the procedure PLF is explained in the program
  FEQ.pas                                                           }

;CONST eps=0.001
;      m=0.075  { mass of the plunger }
;      N=100

{ All the constants are in standard units; meters and Newton/meter;
  a lot of constants are now read out of a datafile                 }

;TYPE real=extended

;VAR  Fpl,x1,x2,G,H1,H2,T,
      xc,yc,l,R,yoff,yon,
      S,d0,k,ystat,yend   :real
;     i                   :integer
;     f,f1,f2             :text
;     state               :string


{ ------- PROCEDURES AND FUNCTIONS ------------------------------ }

;FUNCTION ARCCOS(x: real): real
  ;BEGIN IF x<0 THEN ARCCOS:=ArcTan(Sqrt(1-x*x)/x)+pi
  ;      IF x=0 THEN ARCCOS:=pi/2
  ;      IF x>0 THEN ARCCOS:=ArcTan(Sqrt(1-x*x)/x)
  ;END

{ ------------------------------------------------------------------ }

;FUNCTION force(t: real): real

{ This procedure describes the force applied on the plunger as
  a function of time                                                }

  ;BEGIN IF t<2 THEN force:=0
```

```
                ELSE force:=sin(t*Pi*6)+t-2
  ;END

{ ------------------------------------------------------------ }

;PROCEDURE PLF(var ya: real; var state: string; var Fpl: real)

  ;VAR xa,xb,yb,y,d:real
  ;   FBx,FBy,FBa,FBp,ex,ey:real
  ;   FAa,FAp,FAx,FAy,alpha,beta,h1,h2:real

  ;BEGIN

  {the y-coordinate of A should be chosen in the interval
   (ystat,yend); we now compute the x-coordinate of A              }

   IF (ya<=ystat) AND (ya>=yend)
   THEN
   BEGIN

      xa:=xc-Sqrt(R*R-(ya-yc)*(ya-yc))

      { testing wether the origin is beneath the carrier }
     ;IF state = 'on' THEN y:=yon
                     ELSE y:=yoff
     ;IF ya+Sqrt(xa*xa+ya*ya)*(y-ya)/l>0
      THEN BEGIN y:=yoff; state:='off' END
      ELSE BEGIN y:=yon;  state:='on'  END

     { computation of the coordinates of B }
     ;xb:=xa+S*Sqrt(1-(y-ya)*(y-ya)/(l*l))
     ;yb:=ya+S*(y-ya)/l

     { computation of the force in B }
     ;d:=xb*xb+yb*yb
     ;FBx:=k*(1-d/d0)*xb
     ;FBy:=k*(1-d/d0)*yb

     { its components perpendicular to and along the carrier are
       given by:                                                  }
     ;ex:=(xb-xa)/S
     ;ey:=(yb-ya)/S
     ;FBa:=FBx*ex+FBy*ey
     ;FBp:=ex*FBy-ey*FBx

     { the components of the force in A perpendicular to and along
       the carrier are given by:                                  }
     ;FAa:=FBa
     ;FAp:=S*FBp/l

     { and its (x,y)-coordinates are: }
     ;FAx:=FAa*ex-FAp*ey
```

```
;FAy:=FAa*ey+FAp*ex

{ we now compute the magnitude of the plungerforce which
  with the reaction force compensates the force at A.
  We have to make distinction between four different cases     }
;h1:=xc-xa
;h2:=yc-ya
;IF h1*FAy-h2*FAx>=0
 THEN IF FAx>=0
      THEN BEGIN alpha:=ARCCOS((h1*FAx+h2*FAy)/
                               (Sqrt(h1*h1+h2*h2)
                              *Sqrt(FAx*FAx+FAy*FAy)))
           ;     beta:=ARCCOS(-h2/Sqrt(h1*h1+h2*h2))
           ;     Fpl:=Sin(alpha)
                         *Sqrt(FAx*FAx+FAy*FAy)/
                          Sin(beta)
           ;END
      ELSE BEGIN alpha:=ARCCOS((-h1*FAx-h2*FAy)/
                               (Sqrt(h1*h1+h2*h2)
                              *Sqrt(FAx*FAx+FAy*FAy)))
           ;     beta:=ARCCOS(h2/Sqrt(h1*h1+h2*h2))
           ;     Fpl:=Sin(alpha)
                         *Sqrt(FAx*FAx+FAy*FAy)/
                          Sin(beta)
           ;END
 ELSE IF FAx>=0
      THEN BEGIN alpha:=ARCCOS((h1*FAx+h2*FAy)/
                               (Sqrt(h1*h1+h2*h2)
                              *Sqrt(FAx*FAx+FAy*FAy)))
           ;     beta:=ARCCOS(h2/Sqrt(h1*h1+h2*h2))
           ;     Fpl:=-Sin(alpha)
                          *Sqrt(FAx*FAx+FAy*FAy)/
                           Sin(beta)
           ;END
      ELSE BEGIN alpha:=ARCCOS((-h1*FAx-h2*FAy)/
                               (Sqrt(h1*h1+h2*h2)
                              *Sqrt(FAx*FAx+FAy*FAy)))
           ;     beta:=ARCCOS(-h2/Sqrt(h1*h1+h2*h2))
           ;     Fpl:=-Sin(alpha)
                         *Sqrt(FAx*FAx+FAy*FAy)/
                          Sin(beta)
           ;END

;END
ELSE
BEGIN
    IF ya>ystat THEN ya:=ystat
   ;IF ya<yend  THEN ya:=yend
```

```
      ;Fpl:=force(i*T/N)
      ;H1:=0
      ;G:=0
   ;END
  ;END

{ ------- MAIN PROGRAM ----------------------------------------- }

;BEGIN

{ read data from input file }
 assign(f,'inpdat.dat')
;reset(f)
;readln(f,xc,yc,l,R,yoff,yon,S,d0,k,ystat,yend)
;close(f)

{ initialisation of output files }
;assign(f1,'appforc.dat')
;rewrite(f1)
;assign(f,'displ.dat')
;rewrite(f)
;assign(f2,'position.dat')
;rewrite(f2)

{ initialisation of discretisation: the point t=0 }
;T:=5                    { the total time that we apply the force }
;state:='off'            { at t=0 the switch is off }
;x1:=ystat               { initial position of the plunger }

;writeln(f,0,'  ',x1)
;writeln(f1,0,'  ',force(0))
;writeln(f2,0,'  ',length(state)-2);

;PLF(x1,state,Fpl)
;H1:=(Fpl-force(0))/m    { the acceleration at x1 }
;IF h1>0 THEN h1:=0

{ computations in the next points }
;x2:=x1+0.5*H1*T*T/(N*N)
;PLF(x2,state,Fpl)

;writeln(f,T/N,'  ',x2)
;writeln(f1,T/N,'  ',force(T/N))
;writeln(f2,T/N,'  ',length(state)-2)
```

```
;H2:=(Fpl-force(T/N))/m  { the acceleration at x2 }
;G:=(H1+H2)*T/(2*N)      { the speed at x2 }

;FOR i:=2 to N DO
 BEGIN

       x2:=x1+G*T/N              { computation of x(i*T/N) }
 ;     PLF(x2,state,Fpl)

 ;     writeln(f,i*T/N,'  ',x2);
 ;     writeln(f1,i*T/N,'  ',force(i*T/N));
 ;     writeln(f2,i*T/N,'  ',length(state)-2);

 ;     H2:=(Fpl-force(i*T/N))/m  { acceleration at x(i*T/N) }
 ;     G:=G+(H1+H2)*T/(2*N)      { speed at x(i*T/N) }
 ;     x1:=x2
 ;     H1:=H2

 ;END


{ closing output files }
;close(f)
;close(f1)
;close(f2)

;END.
```

# INEXPLICABLE PAPERBREAKS

# IN

# PAPERMAKING FACTORIES

# Inexplicable Paperbreaks in Papermaking Factories

## Contents

# 1 Introduction

The problem discussed in this report was first given as one of the assignments of the 5-th ECMI modelling week which was held in September 1992 at the Johannes Kepler University in Linz, Austria. This report is not a continuation of the work done in Linz. The approach presented here differs from the one presented in [1] by Grünholz, Haug, Nixon, Omerzu and Pruis.

The problem originates from the paper industry. During one part of the production process the paper tends to break for some unknown reason. The production process of paper can be split in several distinct parts. In the first few parts the paper is actually produced, typically with a speed of 10 to 20 meters per second. At the end of this process the produced paper is rolled on large steel cylinders which can be up to 10 meters wide. After about 1 hour the roll is replaced by an empty one and the full roll is moved to the cutting section of the paper factory. There the roll is unrolled, cut in narrower strips and rolled again on smaller cylinders into smaller rolls.

The breakage of paper can occur during this unrolling and cutting process, typically when there is still between 1 and 3 kilometers of paper on the roll. To avoid the breakage and the subsequent time-consuming removing of the paper from the machines, the unrolling and cutting process is stopped when there is still about 5 kilometers of paper on the roll. This results in a production loss of nearly 10%.

In this report we will look for a possible reason for the paperbreak.

# 2 Problem definition

At some point during the production process of paper, the paper appears to break. The exact cause for this is not known. Some (possible) explanations are enumerated below.

1. Due to high rotational velocities at the beginning of the rolling process, the tangential stresses in the paper can be higher at the inside of the roll than at the outside. The paper can get weak as a result of the large stresses.

2. The curvature of the paper is larger for paper near the steel cylinder than for paper at the outside of the roll. Maybe a large curvature weakens the paper.

3. When the rolling process is stopped too quickly, the stresses in the paper can become too large as a consequence of the large inertia of momentum, thus weakening the paper.

4. The forces applied by the outer layers of the paper roll on the inner layers is larger than the forces applied by the inner layers on the outer layers. The paper of the inner layers may be weaker for this reason. Furthermore, foldings might appear in the inner layers. In the unrolling process the paper can break due to changes in the velocity that result from these foldings.

5. When the paper is rolled on the cylinder, it becomes stressed. During the unrolling process this stress results in some way to the breakage of paper.

6. The stresses at which the paper is put on the roll might not be constant for the whole roll. Maybe the paper breaks where these stresses are largest.

7. The paper might not be homogeneous. This can, however, only be the cause for the problem if this inhomogeneity appears when the paper is already rolled on the cylinder. Two reasons might be that paper at the inner layers suffer from higher temperature gradients, or that water in the inner layers can not vaporize enough.

All the above items are possibilities. It is not clear whether they are reasons for the breakage of the paper. After a closer look of these items, some items seem more likely than others.
Items no. 2, 5, 6 and 7 may be reasons for the paperbreak but it is not clear how. Therefore it is no use to model them even when that was possible.
Item no. 3 is not very likely, because the deceleration time for the roll is very large so that there will not be very large stresses induced during the stopping of the rolling process.
Items no. 1 and 4 (and no. 3) involve the stresses in the paper. The idea is that the paper becomes damaged due to too large stresses in the rolling process. The paper will consequently break in the unrolling/cutting process.

In this report the forces and stresses that act on the paper roll during the rolling process will be modelled. The main question will be whether large

(or negative) forces/stresses in the rolling process can be the reason for the breakage of paper in the unrolling process.

# 3 News about the paper factory

With the aim to get to know more about the sizes of typical paper rolls used in industry, a newspaper factory was phoned.

At the Nederlandse Dagbladunie (NDU) in Rotterdam, they get rolls from a paper factory delivered in any size they want. The rolls they usually work with have the following properties:

| | |
|---|---|
| width: | 1.66 m, |
| weight: | 1100-1200 kg, |
| length: | 15 km, |
| total diameter: | 1.15 m, |
| diameter inner cylinder: | 75 mm, |
| thickness inner cylinder: | 17 mm, |
| material inner cylinder: | cardboard, |
| weight paper per $m^2$: | 45 g, |
| speed of unrolling: | 36-40 km/h. |

After the paper is printed, it is cut and folded. In the process of unrolling and cutting, the paper never breaks. The NDU is insured for breaking and when it should happen, they blame the paper factory that makes the paper for them.

One may notice that the rolls used in the newspaper factory are much smaller than the ones used in the initial problem. Therefore it was decided to phone a paper-making factory: Parenco in Renkum.

The paperrolls they make have the following properties:

| | |
|---|---|
| width: | 8.5 m, |
| weight: | 30000-40000 kg, |
| length: | 70 km, |
| total diameter: | 2.5-3 m, |
| diameter inner cylinder: | 0.5 m, |
| material inner cylinder: | massive iron, |
| speed of rolling: | 75 km/h, |
| speed of unrolling: | 120 km/h. |

As one can see, these rolls are much more alike the ones used in the initial problem. Moreover, at Parenco, the paper sometimes broke. They knew (for sure) that the problem could be found in the rolling process:

Paper comes from the paper-making machine on a belt and is wrapped around a cylinder which is at the end of the belt. The cylinder rotates with a certain speed which is decreased during the process of rolling. (The roll is getting bigger while the paper on the belt arrives at a constant speed.) While the roll is getting full, a second cylinder is brought to its initial speed just above the roll. The paper is cut and with an ingeneous process which involves blowing of air (they worked on it for three years) the loose end is wrapped around the new cylinder. The old roll is moved away and its speed is slowly decreased. The new roll is moved down to position.

This whole process takes about three minutes and several things can go wrong. For example, the air can blow a slug near the edge of the paper and this makes breakages much more likely to occur. In three minutes, the roll will contain about 3 km of paper so this may explain the problem.

The reason given above may apply for paper breakages at Parenco, but they need not be applicable for all paper making factories in the world. Specifically, the initial problem was originated at a factory in Finland and they may not use the ingeneous process of changing rolls as described above, especially since the process was invented at Parenco.

# 4 Introduction to the models

The main questions are, of course, how to model the rolling of the paper and what one should expect from such a model, i.e. which answers are explanations of the breaking of the paper.

A first try could be to model the roll of paper as number of cylinders, making each of them just a bit too narrow, so a certain amount of stress would appear. This is, on second thought, not a very comfortable model, since each cylinder would have a thickness of approximately a millimeter and the total radius would be over a few meters. Thus, one should have thousands of cylinders and this could not be calculated easily anymore, let alone be accurate, because the initial radii of the cylinders are unknown.

One can model the cylinder essentially in two different ways. The first one

could be to model it as a solid, rotating around its axis. The advantage of this is of course, that one can use the linear elasticity theory, based on Hooke's Law. But this can not be an argument to choose this particular model.

The second way to model the paper is to assume it to be a fluid. This may seem strange at first, but it isn't. Paper is not a solid as strong as iron. Thus, it will flow because of the present forces. Therefore one can use the Navier-Stokes Equation to analyze this behaviour.

The solid model is in its simplest form not satisfactory. This is because the paper does in a way behave as a solid, but not the same in all three directions. It would behave in the radial direction with different parameters as in the tangential or axial direction and it will not behave as a linear elastic medium, because of the fact that a too big displacement will deform the paper permanently.

The Newton (fluid) model is also not satisfactory. The difference with the solid model is that now the 'fluid' will only resist to the speed of deformation while in the solid model the solid resists only to the deformation itself. But the paper will not only resist to one and not to the other, but to both.

So a third model is required. Therefore one can introduce the so-called Lodge equation which is a mixture of the Newton and the Hooke model.

By now one can introduce three different models. The first one will be based on Hooke's Law, where displacements are only temporarily and the initial position will be regained after the pressure is lifted. The second one will be a Newton Fluid, where the displacement will be permanent and an actual flow will be present and the third one is an easy mixture of these, based on the Lodge equation or the so-called Maxwell models.

The other question is, what one wants for an answer. For instance, if the tangential stress in the cylinder is at a particular place negative, there could be folding of the paper which in turn can result in damage. Another answer could be that, if the radial or tangential stress will become too large, the paper will also be damaged and thus break in the unrolling process.

One thing that can be said for all models is that the increase in time of $R_{\max}$, the radius of the cylinder, can be neglected, since the speed at which it grows is approximately a few meters per hour, while the rotation of the cylinder is of the order turns per second. Thus, we do not consider the increase in time of the radius and take it to be a constant.

By neglecting the increase in time and thinking only of a stationary situation, we will not have to bother with time-derivatives etc. This is a major advantage.

# 5 The solid state model

The paper roll is considered as a massive solid cylinder. Further assumptions are:

- The problem is 2-dimensional, i.e. there is no displacement in the axial direction, and all used variables are independent of the axial coordinate.
- The problem is rotation symmetric.
- Time is not considered (stationary situation).
- The paper on the cylinder is considered to be of linear elastic and isotropic material. This is a major simplification, since paper is in fact not isotropic.
- The influence of gravity is neglected.

As a result of these assumptions, one can write the displacement $u$ of a particle in the radial direction as a function of the radius $r$: $u = u(r)$. Besides the displacement one also considers the deformation $e$. The deformation tensor is then given by (see [2])

$$\begin{aligned} e_{rr} &= \frac{\mathrm{d}u}{\mathrm{d}r}, \\ e_{\theta\theta} &= \frac{u}{r}, \\ e_{r\theta} &= e_{rz} = e_{\theta r} = e_{\theta z} = e_{zr} = e_{z\theta} = e_{zz} = 0. \end{aligned}$$

Since one assumed that the material is linear elastic and isotropic, Hooke's law may be applied:

$$\sigma_{ij} = \lambda \delta_{ij} e_{kk} + 2\mu e_{ij},$$

where $\sigma_{ij}$ is the tension in the $j$-direction on a plane $i$ = constant, and $\lambda$ and $\mu$ are the so-called Lamé parameters, that relate to the Young's modulus $E$ and Poisson ratio $\nu$ as follows:

$$\begin{aligned} E &= \frac{3\mu(\lambda+2\mu)}{\lambda+3\mu}, \\ \nu &= \frac{\lambda}{2(\lambda+3\mu)}. \end{aligned}$$

Using the relations for the deformation tensor, one obtains

$$\begin{aligned} \sigma_{rr} &= (\lambda+2\mu)\frac{\mathrm{d}u}{\mathrm{d}r} + \lambda\frac{u}{r}, \\ \sigma_{\theta\theta} &= \lambda\frac{\mathrm{d}u}{\mathrm{d}r} + (\lambda+2\mu)\frac{u}{r}, \\ \sigma_{zz} &= \lambda(\frac{\mathrm{d}u}{\mathrm{d}r} + \frac{u}{r}), \\ \sigma_{r\theta} &= \sigma_{rz} = \sigma_{\theta r} = \sigma_{\theta z} = \sigma_{zr} = \sigma_{z\theta} = 0. \end{aligned}$$

Next, to compute the tensions $\sigma_{rr}$ and $\sigma_{\theta\theta}$ one considers equilibrium of forces. In this equilibrium, a term for centrifugal force is included, since the paper roll is rolling very fast. The term for the centrifugal force is $\rho\omega^2 r$, where $\rho$ is the mass density of the paper, and $\omega$ is the angular velocity of the roll. Equilibrium of forces gives:

$$\frac{\mathrm{d}\sigma_{rr}}{\mathrm{d}r} + \frac{1}{r}(\sigma_{rr} - \sigma_{\theta\theta}) + \rho\omega^2 r = 0.$$

This leads to a differential equation for the displacement

$$(\lambda+2\mu)\left\{\frac{\mathrm{d}^2u}{\mathrm{d}r^2} + \frac{1}{r}\frac{\mathrm{d}u}{\mathrm{d}r} - \frac{1}{r^2}u\right\} + \rho\omega^2 r = 0,$$

that can be solved to give

$$u(r) = Ar + \frac{B}{r} - \frac{\rho\omega^2 r^3}{8(\lambda + 2\mu)}, \qquad R_{\min} \leq r \leq R_{\max}.$$

$A$ and $B$ are constants that may be found using boundary conditions:

$$\sigma_{rr}(R_{\max}) = 0,$$

i.e. there is no radial tension on the outside of the cylinder, and

$$u(R_{\min}) = 0,$$

i.e. there is no displacement at the inner cylinder. The tensions in the radial and in the tangential direction are easily found:

$$\begin{aligned}
\sigma_{rr} &= 2A(\lambda + \mu) - 2\mu\frac{B}{r^2} - \frac{2\lambda + 3\mu}{4(\lambda + 2\mu)}\rho\omega^2 r^2, \qquad R_{\min} \leq r \leq R_{\max}, \\
\sigma_{\theta\theta} &= 2A(\lambda + \mu) + 2\mu\frac{B}{r^2} - \frac{2\lambda + \mu}{4(\lambda + 2\mu)}\rho\omega^2 r^2, \qquad R_{\min} \leq r \leq R_{\max}.
\end{aligned}$$

One is mainly interested in the tension $\sigma_{\theta\theta}$ since if this tension is too large it may cause damage in the paper. It can be seen that for small radius $r$ this tension will indeed become large due to the term with $\frac{1}{r^2}$.

# 6 The fluid state model

## 6.1 Newton fluid

One starts with the conservation of momentum :

$$\rho\dot{\vec{v}} = \rho\vec{f} + \vec{\nabla} \cdot \sigma, \tag{1}$$

where the dot denotes the material derivative, $\rho$ stands for the density, $\vec{f}$ for the body forces and $\sigma$ for the stress tensor.

Introducing $\sigma$ as

$$\sigma = -p\mathbf{I} + 2\eta\mathbf{D},$$

this results in the very famous Navier Stokes Equations (see [3]):

$$\dot{\vec{v}} = -\frac{1}{\rho}\vec{\nabla}p + \nu\nabla^2\vec{v} + \vec{f},$$

where $\nu$ is $\frac{\eta}{\rho}$.

Rewriting these in cylindrical coordinates, neglecting the body forces, taking

$$v_\theta = v_\theta(r), \quad v_r = 0, \quad v_z = 0, \quad p = p(r),$$

and only considering a stationary situation results in

$$\frac{v_\theta^2}{r} = \frac{1}{\rho}\frac{\mathrm{d}p}{\mathrm{d}r},$$

and

$$0 = \nu\left(\frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}r}\left(r\frac{\mathrm{d}v_\theta}{\mathrm{d}r}\right) - \frac{v_\theta}{r^2}\right).$$

The solution of these equations is given by

$$v_\theta(r) = Ar + \frac{B}{r},$$

and

$$p(r) = p_0 + \rho\left(\frac{1}{2}Ar^2 - \frac{B^2}{4r^4}\right).$$

where $p_0$ is the atmospherical pressure and $A$ and $B$ are unknown constants.

The remaining nonzero stress component is

$$\sigma_{r\theta} = -\mu \frac{3B}{r^2},$$

where $\mu$ is an unknown constant.

From the pressure and the stress-term it is likely that $B$ is negative. So both the $\sigma_{r\theta}$ stress and the pressure $p$ can increase rapidly when the radius of the inner cylinder (on which the paper is rolled) decreases. Thus this model suggests that the stresses near the steel cylinder reaches a certain maximum so that the paper can be damaged. The exact value of the stresses can only be estimated.

## 6.2 Maxwell model

Again, model the roll of paper as a fluid. The velocity of a particle in the roll can be expressed by the law of conservation of momentum, equation (1). This equation can be solved when $\sigma$ is known.

To express the non-elasticity of paper, the following Lodge's equation is appropriate to use for $\sigma$ (according to [4]):

$$\sigma = G \int_{-\infty}^{t} \frac{1}{\lambda} e^{(t'-t)/\lambda} \mathbf{B}(t, t') dt'.$$

Here, $\lambda$ is the relaxation time and $G$ is the stiffness of the fluid. The Finger elasticity tensor $\mathbf{B}$ is a measure for the displacement of a particle in $[t', t]$.

Lodge's equation can be written in differential form, which is called the (upper-convected) Maxwell equation (see [4]):

$$\dot{\sigma} - (\vec{\nabla}\vec{v})^T \cdot \sigma - \sigma \cdot (\vec{\nabla}\vec{v}) + \frac{1}{\lambda}\sigma = \frac{G}{\lambda}\mathbf{I} \qquad (2)$$

Equations (1) and (2) can be written into cylindrical coordinates and can be solved. The following assumptions have been made:

- A stationary situation is assumed. This is allowed because while the roll makes a whole turn, it only becomes one slice of paper bigger so there doesn't change much. This implies:

$$\frac{\partial \vec{v}}{\partial t} = 0 \quad \text{and} \quad \frac{\partial \sigma}{\partial t} = 0.$$

  Moreover it is assumed that $\dot{\sigma} = 0$.

- Body forces are neglected: $\vec{f} = 0$.

- It is assumed that there are no velocities in the $z$-direction, or $v_z = 0$. Moreover, the displacement of a particle in the $r$-direction will be very small, compared with the displacement in the $\theta$-direction: the roll will make a whole turn while it becomes only a slice of paper bigger. Therefore, $v_r$ is neglected. Also, from axis-symmetry, $v_\theta = v_\theta(r)$.

After (2) has been rewritten into cylindrical coordinates, it can be partially solved:

$$\sigma = A \cdot \begin{pmatrix} 2\lambda\frac{v_\theta}{r} & 1 & \sigma_{rz} \\ 1 & 2\lambda\frac{\mathrm{d}v_\theta}{\mathrm{d}r} & \sigma_{\theta z} \\ \sigma_{zr} & \sigma_{z\theta} & 0 \end{pmatrix} + G \cdot \begin{pmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

where,

$$A = \left(\frac{\mathrm{d}v_\theta}{\mathrm{d}r} - \frac{v_\theta}{r}\right) \frac{\lambda G \cos(\theta)}{1 - 4B},$$

and

$$B = -\lambda^2 \frac{\mathrm{d}v_\theta}{\mathrm{d}r} \frac{v_\theta}{r}.$$

Also, one of the following two cases is true:

- $\sigma_{zr} = \sigma_{z\theta} = \sigma_{rz} = \sigma_{\theta z} = 0$, or
- $B = 1$.

In the first case, $\sigma$ can be solved and with this, (1) can be solved. The resulting equations are rather large and therefore omitted from this paper.
In the last case, $B = -\lambda^2 \frac{\mathrm{d}v_\theta}{\mathrm{d}r}\frac{v_\theta}{r} = 1$ hence $v_\theta(r)$ can be solved. (As an initial condition, $v_\theta$ is to be taken equal to $v_\theta(r_0)$ on $r = r_0$ where $r_0$ is an arbitrary radius).

$$v_\theta(r) = \sqrt{v_\theta^2(r_0) - \frac{r^2 - r_0^2}{\lambda^2}}.$$

From this, $\sigma_{\theta\theta}$ can be solved as a function of $r$:

$$\sigma_{\theta\theta} = \left(\frac{1}{\lambda^2 v_\theta^2(r_0) - r^2 + r_0^2} + \frac{1}{r} + \frac{3}{2}\right) \cdot G\cos(\theta).$$

# 7 Conclusions and recommendations

## Conclusions

The actual cause why paper breaks is not known and is very difficult to determine. In fact, it is not even known whether the paper breaks during rolling, unrolling or during cutting. A deterministic model for paperbreaks is therefore impossible at this moment.

A possible reason for the breakage of paper might be that the paper gets weak during the rolling process due to large (or negative) tangential stresses. The rolling process can both be modelled with solid state and with fluid state models. It is difficult, however, to take into account the specific characteristics of paper, e.g. elasticity, anisotropy and plasticity. Moreover, typical paper parameters such as Young's modulus, Poisson's ratio and the frictioncoefficient are hard to find in literature. Parameters for the fluid state model (e.g. viscosity and the relaxation time) must be estimated.

The simple models discussed in this report all showed that the tangential stresses increase when the distance to the axis of the cylinder becomes smaller. It seems possible that the paper gets damaged near the inside of the paperrol due to these stresses.

## Recommendations

The models discussed in this report are very simple ones. By finding better constitutive relations for the stresses in this geometry the models might be improved.

After useful parameters for paper are found, calculations that will give a more quantitative analysis of the problem of paperbreaks can be performed. Especially the critical stress at which (damaged) paper breaks has to be seeked. The fact that the radius of the cylinder grows in time must be included in the calculations.

The problem of paper breaks might (for now) be solved by using larger cylinders to roll the paper on. The stresses in the paper should, according to our models, become smaller. This is however an expensive solution. The cylinders itself will not only be more expensive, but they will also be heavier. Stronger machines are needed then.

# References

[1] T. Grünholz, E. Haug, C. Nixon, A. Omerzu and G. Pruis,
*Breaking in Paper Rolls*,
report of the 5-th ECMI-modelling week in Linz, 1992 (not published).

[2] A. van der Ven,
*Continuümstheorie*,
lecture notes of the Department of Mathematics and Computer Science,
Technical University of Eindhoven, 1992/1993 (in Dutch).

[3] G. Vossers,
*Fysische Transportverschijnselen voor W*,
lecture notes of the Department of Technical Physics,
Technical University of Eindhoven, September 1986 (in Dutch).

[4] H. Meijer, G. Peters and A. Spoelstra,
*Polymeerverwerking en Reologie; Deel I: Basis*,
lecture notes of the Department of Mechanical Engineering,
Technical University of Eindhoven, June 1992 (in Dutch).

# WASHING DISHES

## A simple model to compare detergents

# Contents

# 1 Problem Description

In an advertisement of a well-known detergent it is posed that with this detergent $x\%$ more dishes can be cleaned than with every other product. The question is: Is this reasonable?

# 2 The Mathematical Model

## 2.1 Introduction

For comparing two detergents we calculate how many dishes we can clean per volume unit detergent if we are cleaning under realistic circumstances. As a result the temperature of the water plays an important role.

## 2.2 Basic Assumptions

What is modelled is the cleaning under circumstances as realistic as possible (and as simple as possible) with some constants dependent on the kind of used detergent. Therefore the following assumptions are made:

- The used water has a "normal" starting temperature (f.i. $60°C$).
- The used amount of water is "normal" (f.i. 10 litres).
- The used amount of detergent is "normal". It's assumed that this amount is given by the package of the detergent (f.i. 1 teaspoon).
- All dishes are equally dirty before the cleaning (f.i. 1 *mmol grease/dish*).
- All dishes are totally clean after washing (e.g. 0 *mmol grease/dish*).
- The model is 0-dimensional (in space). This is allowed because during the washing up the water is mixed. That means that the water is homogenuous.
- The person who's washing has to stop when there is not enough soap left in the water or when the temperature of the water is too low.

## 2.3 Continuous Parameters

There are three continuous parameters:

- Temperature

  During the washing up the water cools down naturally. In this model the speed with which this happens is linear with respect to the difference between the temperature of the water and the temperature of the surrounding area.

- Velocity

  In this model the speed with which the person is cleaning is constant in time. It does not depend on concentration or temperature. We assume the person is not getting tired.

- Concentration

  To handle with the concentration we must look at the washing up at molecule level. There are three forms in which a soap molecule can be in the water:

  - free and active
  - free but inactive
    Free soap molecules can be inactive, because they need a certain energy level to be able to interact with dirt molecules.
  - bounded to dirt

  We are only interested in free and active soap molecules.

## 2.4 The Formulas

The used notation is explained in appendix A.
With the use of the description of the continuous parameters we define the following formulas to describe the process:

$$T(t) = T_a + (T(0) - T_a)e^{-kt} \tag{1}$$

$$v(t) = v \tag{2}$$

$$A(t) = tv \tag{3}$$

$$c(t) = c(0) - svdt - \alpha(T(t))c(0) \tag{4}$$

$$\alpha(T(t)) = \begin{cases} 0 & for\ T(t) \geq T^* \\ \dfrac{T^* - T(t)}{T^* - T_i} & for\ T(t) \leq T^* \end{cases} \tag{5}$$

Explanation for the formula of the concentration:
At the beginning the amount of soap is $c(0)$. During the washing up the amount of dirt that is coming in per time unit is $dv$. To bind these dirt molecules we need $sdv$ soap molecules per time unit.
Soap can become inactive when the temperature of the water is lower than $T^*$. The fraction that becomes inactive is $\alpha(T(t))c(t)$. Because this is too difficult to calculate (analitically) we try as a first approximation $\alpha(T(t))c(0)$, with $\alpha$ linear from 0 to 1 when the temperature goes down from $T^*$ to $T_i$. This seems to be a very rough approximation. But even with $c(t)$ in this term, it's impossible to get a better approximation, because we know nothing about $\alpha$.

To get an idea about the shape of the parameters we give an example belonging to formula (1), (4) and (5). The shapes of the functions are shown in figures 1 to 3. The values used to obtain these pictures are given in appendix B.

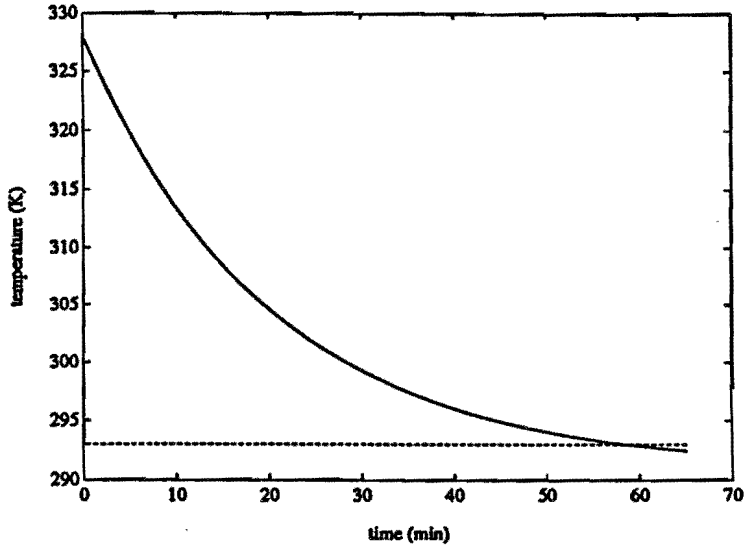



Figure 1: Temperature as a function of time

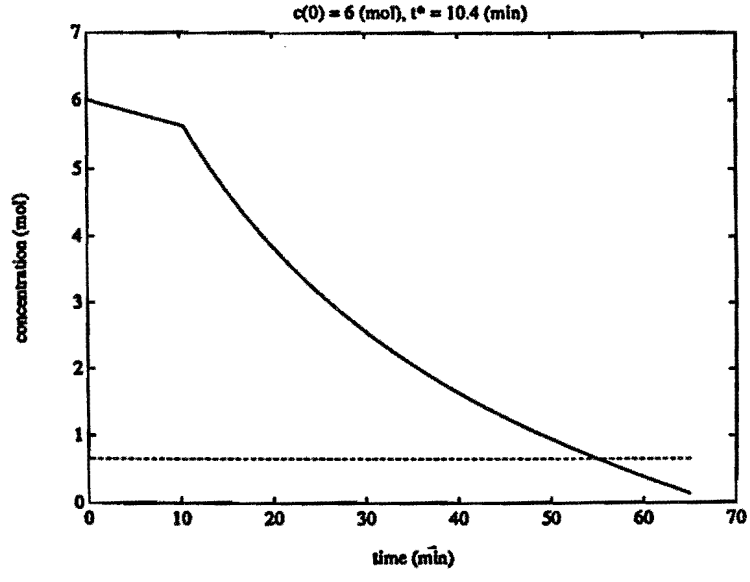



Figure 2: Concentration as a function of time

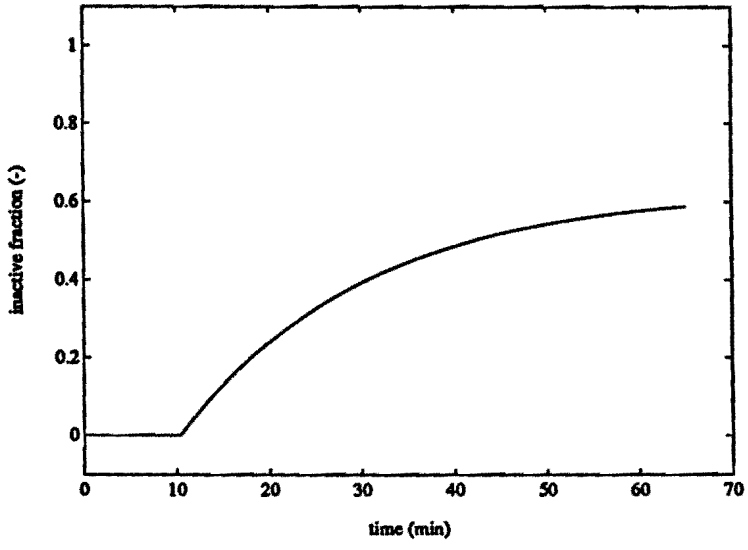



Figure 3: Inactive fraction as a function of time

# 3 Calculations

## 3.1 General Calculations

If we take $e^{-kt_s} = 0$ ($t_s$ is rather big) we find the following formulas:

$$t_s \simeq \frac{c(0)\left(\frac{T_a - T_i}{T^* - T_i}\right) - c_s}{svd} \tag{6}$$

$$t_{sT} = \frac{-1}{k} \ln\left(\frac{T_s - T_a}{T(0) - T_a}\right) \tag{7}$$

$$\frac{A(t_s)}{a} \simeq \frac{c(0)\left(\frac{T_a - T_i}{T^* - T_i}\right) - c_s}{sda} \tag{8}$$

$\frac{A(t_s)}{a}$ is the quantity we use to compare two detergents.
If we have to stop cleaning because the temperature is too low, the amount of dishes we can clean is $A(t_{sT})$. Because the speed is independent of the used detergent, $A(t_{sT})$ is the maximum amount that can be cleaned. Result:
If we can clean $A(t_s)$ dishes with detergent 1, it is only possible to clean $(1+x)A(t_s)$ with detergent 2 if

$$(1+x)A(t_s) \leq A(t_{sT}) \tag{9}$$

Suppose (9) holds. Then we can try to clean $x$% more by using detergent 2. What can change is: $c(0)$, $s$ or $T^*$. In the following sections we discuss these three possibilities.

## 3.2 The Influence of $c(0)$

Assumptions:
$\frac{c_1(0)}{a_1} \neq \frac{c_2(0)}{a_2}$, $s_1 = s_2$ and $T_1^* = T_2^*$.
We assume that $a_1$, $a_2$ and $c_1(0)$ are fixed and we try to find a $c_2(0)$ such that we can clean $x\%$ more. We find the following formula:

$$\frac{c_2(0)}{a_2} = (1+x)\frac{c_1(0)}{a_1} + \frac{c_s}{a_2}\left[1 - \frac{a_2}{a_1}(1+x)\right]\left(\frac{T^* - T_i}{T_a - T_i}\right) \qquad (10)$$

We can write this as: $\frac{c_2(0)}{a_2} = (1+x)\frac{c_1(0)}{a_1} + \epsilon$ where $\epsilon > 0$ and relatively small.
In words:
To clean $x\%$ more dishes the concentration of soap molecules must be not only $x\%$ higher but you need some extra soap. That sounds reasonable, because the speed is constant. So, cleaning $x\%$ more means cleaning for a longer time. During this extra time the temperature goes down and therefore more soap becomes inactive. That's why we need more than $x\%$ extra soap.

Figure 4 shows how $c(t)$ changes if the starting concentration is changed. $c_2(t)$ is the dotted line (also in figure 5 and 6).

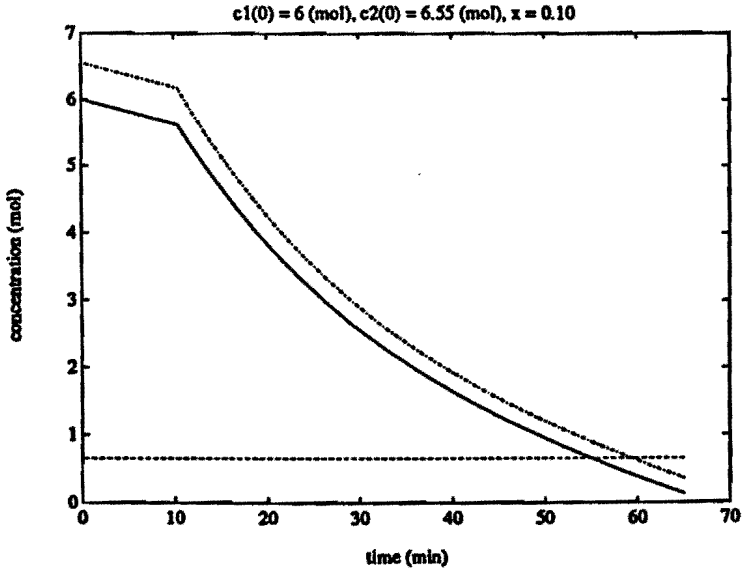



Figure 4: concentration as a function of time; $\frac{c_1(0)}{a_1} \neq \frac{c_2(0)}{a_2}$

## 3.3 The Influence of $s$

Assumptions:
$\frac{c_1(0)}{a_1} = \frac{c_2(0)}{a_2}$, $s_1 \neq s_2$ and $T_1^* = T_2^*$.
Analogously as in the previous section we find the following formula:

$$s_2 = \frac{\left(\frac{c_1(0)}{a_1}\left(\frac{T_a - T_i}{T^* - T_i}\right) - \frac{c_s}{a_2}\right)}{\left(\frac{c_1(0)}{a_1}\left(\frac{T_a - T_i}{T^* - T_i}\right) - \frac{c_s}{a_1}\right)} \frac{s_1}{1+x} \tag{11}$$

Because $c_s$ is very small compared with $c_1(0)$ we can write this as: $s_2 \simeq \frac{s_1}{1+x}$.
In words:
To clean $x$% more dishes with detergent 2, $s_2$ must be $x$% smaller. That sounds reasonable, because then we need less soap to bind the dirt. The value of $s$ is a characteristic of the used type of soap molecules. Therefore decreasing $s$ means that "better" soap molecules must be used.

Figure 5 shows how $c(t)$ changes if $s$ is changed.

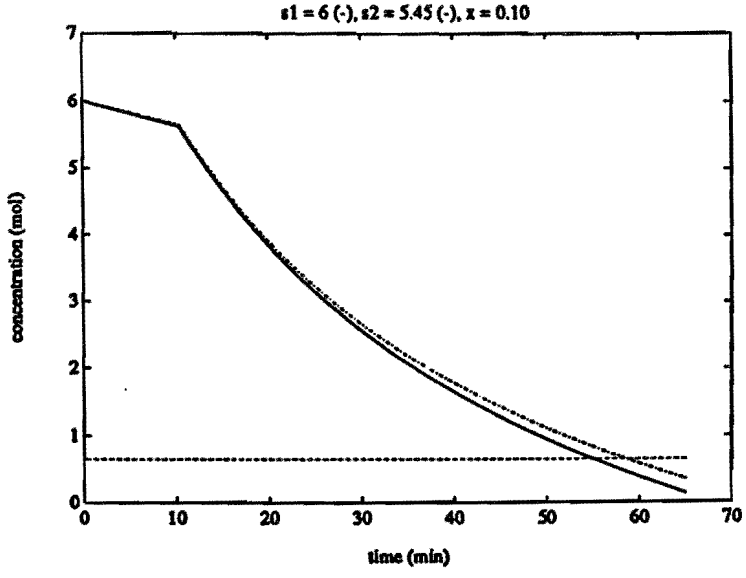



Figure 5: concentration as a function of time; $s_1 \neq s_2$

## 3.4 The Influence of $T^*$

Assumptions:
$\frac{c_1(0)}{a_1} = \frac{c_2(0)}{a_2}$, $s_1 = s_2$ and $T_1^* \neq T_2^*$.
Analogously as in the previous section we find the following formula:

$$T_2^* = T_i + \frac{T_1^* - T_i}{\frac{c_s}{c_2(0)}\left(\frac{T_1^* - T_i}{T_a - T_i}\right) + (1+x)\left(1 - \left(\frac{T_1^* - T_i}{T_a - T_i}\right)\frac{a_2 c_s}{a_1 c_2(0)}\right)} \tag{12}$$

Because $c_s$ is very small compared with $c_2(0)$ we can write this as:
$T_2^* \simeq T_i + \frac{T_1^* - T_i}{1+x} = \frac{T_1^* + xT_i}{1+x}$.
In words:
To clean $x$% more dishes with detergent 2, $T_2^*$ must be less than $x$% smaller. It sounds reasonable that $T_2^*$ must be smaller than $T_1^*$, because then all the soap molecules remain active for a longer time. $T_2^*$ must be less than $x$% smaller means that the influence of $T^*$ is rather big.
$T^*$ is also a characteristic of the used type of soap molecules.

Figure 6 shows how $c(t)$ changes if $T^*$ is changed.

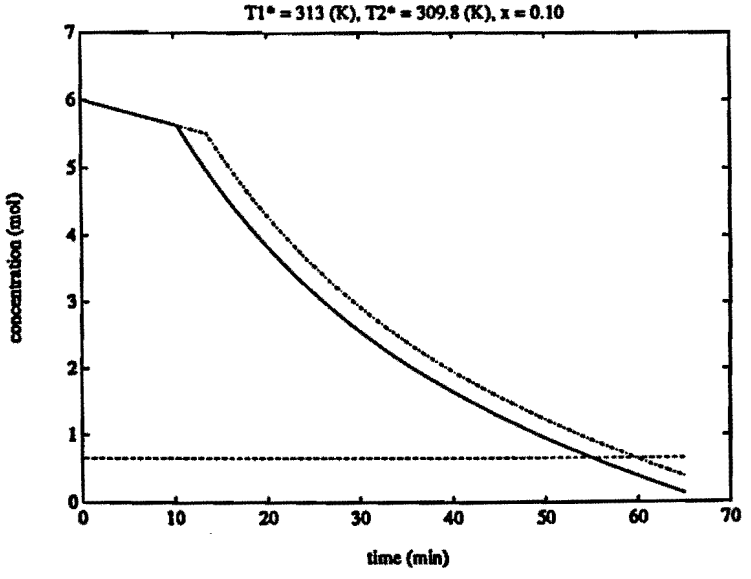



Figure 6: concentration as a function of time; $T_1^* \neq T_2^*$

# 4 Conclusions

Further analyses must be done to get more accurate answers. With the very simple model we have used the following conclusions can be made:

- Cleaning $x\%$ more by using another detergent is only possible when the temperature is not a restricting factor.

- If temperature is not a restricting factor there are two ways to improve a detergent:
  - use "better" soap molecules. "Better" means: also active at a lower temperature or less molecules can solve a dirt molecule.
  - put more soap molecules in the detergent.

# A Notation

| name | symbol | unit |
|---|---|---|
| Amount of used detergent | $a_1$, $a_2$ | $ml$ |
| Time | $t$ | $min$ |
| Stopping time concerning the concentration | $t_s$ | $min$ |
| Stopping time concerning the temperature | $t_{sT}$ | $min$ |
| Amount of dirt per dish | $d$ | $mol$ |
| Percentage to clean more with Detergent 2 | $x$ | $-$ |
| Temperature at time $t$ | $T(t)$ | $K$ |
| Temperature of the area | $T_a$ | $K$ |
| Temperature where all soap is inactive | $T_i$ | $K$ |
| Temperature where soap starts becoming inactive | $T_1^*$, $T_2^*$ | $K$ |
| Stopping temperature | $T_s$ | $K$ |
| Temperature decreasing factor | $k$ | $\frac{1}{min}$ |
| Speed of cleaning dishes | $v$ | $\frac{1}{min}$ |
| Amount of cleaned dishes at time $t$ | $A(t)$ | $-$ |
| Number of soap molecules necessary to bind 1 dirt molecule | $s_1, s_2$ | $-$ |
| Amount of active soap molecules at time $t$ | $c(t)$ | $mol$ |
| Amount of active soap molecules at time 0 | $c_1(0)$, $c_2(0)$ | $mol$ |
| Stopping concentration | $c_s$ | $mol$ |
| Fraction of inactive and free soap at time $t$ | $\alpha(T(t))$ | $-$ |

Remark: subindex 1 or 2 means that the quantity depends on the kind of detergent.

# B Used values in the example

For the figures 1 to 3 we have used the following values:

| parameter | value | unit |
|---|---|---|
| $a$ | 2 | $ml$ |
| $d$ | $1 \cdot 10^{-3}$ | $mol$ |
| $x$ | 0.1 | – |
| $T_a$ | 291 | $K$ |
| $T_i$ | 278 | $K$ |
| $T^*$ | 313 | $K$ |
| $T_s$ | 293 | $K$ |
| $k$ | 0.05 | $1/min$ |
| $v$ | 6 | $1/min$ |
| $s$ | 6 | – |
| $c(0)$ | 6 | $mol$ |
| $c_s$ | 0.65 | $mol$ |

For the figures 4 to 6 we have used the following extra values:

| parameter | value | unit |
|---|---|---|
| $a_2$ | 2 | $ml$ |
| $c_2(0)$ | 6.55 | $mol$ |
| $s_2$ | 5.45 | – |
| $T_2^*$ | 309.8 | $K$ |

# EXPERIMENTAL DESIGN

# AND

# QUALITY-LOSS FUNCTION

## Abstract

The paper "Experimental Design and Quality-Loss Function" is studied and discussed. Its subject is the minimization of a quality-loss function – or cost function – if the parameters on which its value depends are unknown.

The way the authors performed the modelling of a "real-life" problem is studied, as well as whether or not the assumptions are realistic, and if the mathematical methods used are valid.

Because of a rather vague description of the "real-life" problem considered, and as a consequence of the lack of numerical data, the modelling itself can not be judged. Some of the assumptions made, and a few of the mathematical techniques used, however, seem to be questionable. They are studied, and alternatives are presented if possible.

# Experimental Design and Quality-Loss Function

## Introduction

The subject of this report is "inverse modelling". Instead of considering a "real-life" problem and making a corresponding mathematical model, we were given a paper in which such a model was readily presented. The question is to find out how the authors of the paper modelled the problem under consideration, what assumptions they made, and what mathematical methods they used.

The paper we will study in this report is called "Experimental Design and Quality-Loss Function". Its subject is the minimization of a quality-loss function – or cost function – if the parameters on which its value depends are stochastic.
Unfortunately, the authors of the paper have only briefly described the "real-life" problem they modelled. It is merely presented as an example – and, as a matter of fact, the minimization methods derived seem to be applicable in far more cases.
As a consequence, we will not go into details as far as the modelling is concerned. We will focus on the assumptions made – checking whether or not they seem to be realistic in the mentioned "real-life" problem – and their mathematical importance. The mathematical methods used will also be criticised.

The outline of this report is as follows.
In the first section, the paper is described and analysed – a process which we called "inverse modelling". The description is a global summary of the paper; its analysis yields a list of assumptions and/or questionable mathematical statements and methods.
These assumptions and questionable statements are considered in the second section of this report. If assumptions appear not to be realistic indeed, or if we do not trust the mathematics, we try to find appropriate alternatives.

# 1 "Inverse Modelling": Description and Analysis of the Paper

## 1.1 Description of the Paper

The paper deals with a fairly complicated problem, which actually consists of two parts. At first, regression analysis is applied to find out how input parameters of a process – e.g. chemicals in photographic films – and the corresponding output parameters – e.g. the quality of the colors – are related. This yields a so-called response function.
Once the response function is known, it is principally possible to predict the output variables if the input is known. It happens that optimal (or desired) values for the output variables are known. A *quality-loss function* is introduced, describing for each set of input parameters how far the corresponding output parameters are from being optimal.
The problem addressed in the paper is to minimize the quality-loss function. This would be quite easy if there were no additional problem – but there is. In practice, it appears to be impossible to provide for an exact set of input parameters; one will have to do with an approximation.
The question is how to "minimize" the quality-loss function, even if the input variables are not exactly known.

The paper consists of two main parts. The first part deals with the mentioned regression analysis, and in the second part the problem of minimizing the quality-loss function is addressed. We will maintain this division throughout this report.

### 1.1.1 The Regression Analysis Part

The aim of the first part of the paper is to find a relation between certain input variables $x_1, \ldots, x_k$ and certain responses $y_1, \ldots, y_n$. A possible interpretation would be that the $x_i$ represent amounts of chemicals in a photographic film, and that the responses $y_i$ correspond to some quality parameters of the resulting pictures.

To determine the relation between input and output parameters, several experiments are conducted, with different inputs. For each of the $N$ experiments, input and response are measured. Using these measurements, the authors try to find a mathematical formula expressing the response as a function of the input variables:

$$\mathbf{y} = f(\mathbf{x}) + \mathbf{e} \tag{1}$$

where $\mathbf{x}$ is a $k$-dimensional vector of input variables, $\mathbf{y}$ is an $n$-dimensional vector

of responses and **e** is an error vector.

The authors propose to determine the response functions using quadratic regression, i.e. to assume that the functions look like

$$y_r = \beta_{0r} + \sum_{i=1}^{k} \beta_{ir} x_i + \sum_{i,j=1}^{k} \beta_{ijr} x_i x_j + e_r. \qquad r = 1, \ldots, n \tag{2}$$

The optimal values for the coefficients $\beta_i$ and $\beta_{ij}$ can be found by the method of least squares. We will not go into detail as far as this method is concerned; details can be found in the paper itself. Intuitively spoken, the method yields values for the coefficients such that the sum of the squared lengths of the *vertical* deviations – as displayed in Figure 1 – is minimized ($k = 1, n = 1$).

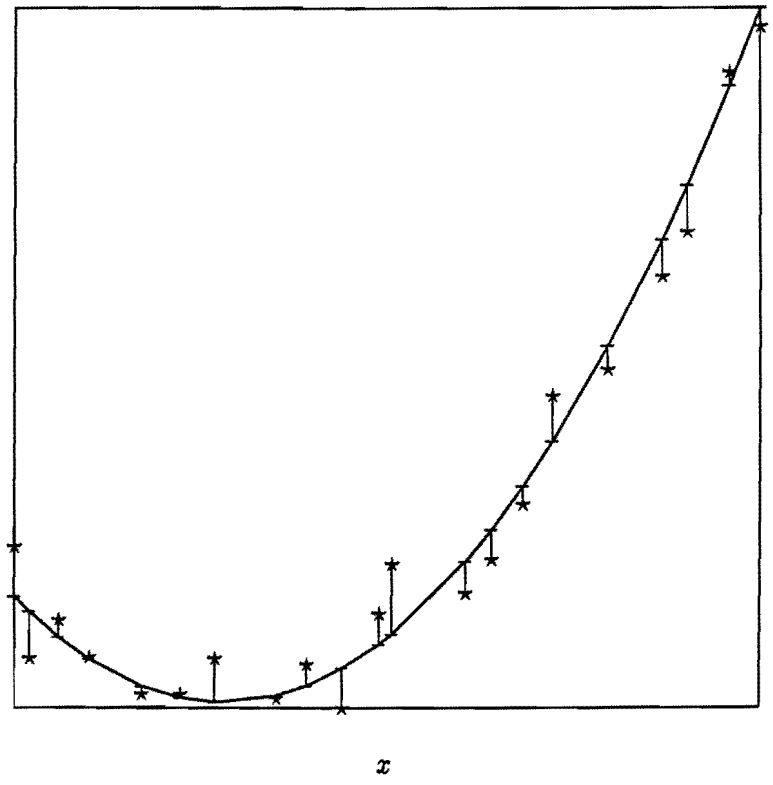



Figure 1: Quadratic regression ($k = 1, n = 1$). The sum of the squared lengths of the *vertical* deviations is minimized.

### 1.1.2 The Minimization Part

In the second part of the paper, the authors assume the response vector $\mathbf{y} = (y_1, y_2, \ldots, y_n)^T$ to be known exactly if the input vector $\mathbf{x} = (x_1, x_2, \ldots, x_k)^T$ is given. The response function is assumed to be quadratic, and – which is more important – the values of $\beta_i$ and $\beta_{ij}$ are now considered to be exact instead of regression results. Thus,

$$y_r = \beta_{0r} + \sum_{i=1}^{k} \beta_{ir} x_i + \sum_{i \leq j} \beta_{ijr} x_i x_j \qquad r = 1, \ldots, n \tag{3}$$

Furthermore, the authors assume optimal values for each response $y_r$ to be known; they will be denoted by $t_r$.

Their aim is to choose input parameters $x_i$ such that the value of a target function – which they call *quality-loss function* – is minimized. If the $x_i$ denote amounts of chemicals in a photographic film, and the $y_i$ describe the quality of the resulting pictures, this quality-loss function gives a mathematical expression for how far the picture quality is apart from the desired quality.
However, a problem arises. Even if it is possible to determine the analytical minimum of the quality-loss function, say $\mathbf{x}_0$, it is practically impossible to actually use this vector as input vector. It may happen that $\mathbf{x}_{0,i} = \sqrt{2}$ – and not even the most precise machine will ever be able to measure the exact amount of $\sqrt{2}$ grams, pints, pounds, ... !
Thus, the actual input will be $\mathbf{x}_0 + \mathbf{z}$, $\mathbf{z}$ denoting some stochastical error.
The authors of the paper now wish to find the best "operating minimum" of the quality-loss function. They want to know for which value of $\mathbf{x}_0$ the average value of the quality-loss function – considering all possible actual inputs $\mathbf{x}_0 + \mathbf{z}$ – is minimal. The distribution of $\mathbf{z}$ is given.

The authors propose the quality-loss function

$$L(\mathbf{y}) = \sum_{r=1}^{n} \{(y_r - t_r) \cdot w_r\}^2. \tag{4}$$

The $w_r$ represent *weight factors* that are large if the corresponding $y_r$ is considered to be important. For instance, as Kodak makes lots of advertisements saying that "Kodak colors are the best", and if $y_1$ expresses the color quality of a film, $w_1$ will be very large in the Kodak quality-loss function.
Since each $\mathbf{y}$ is known as a function of $\mathbf{x}$, this loss function may also be written as a function of $\mathbf{x}$.

$$QL(\mathbf{x}) = L(\mathbf{y}(\mathbf{x})). \tag{5}$$

Note that since each $y_r$ is a quadratic function of $x_1, \ldots, x_k$, the function $QL$ is a polynomial of degree 4 in $x_1, \ldots, x_k$

How to find the best *operating minimum* of the quality-loss function ? The paper suggests the following method.
First determine all *local* minima of $QL$. Since $QL$ is a polynomial of degree 4, there are at most 3 extremes, so at most 2 local minima. A choice will be made between these minima to find the best operating minimum.
Let $\mathbf{x}_0$ be such a local minimum. Assume that instead of $\mathbf{x}_0$, $\mathbf{x} = \mathbf{x}_0 + \mathbf{z}$ is input, the vector $\mathbf{z}$ representing the stochastic error. Instead of choosing that input $\mathbf{x}_0$ where $QL(\mathbf{x}_0)$ is minimal (the global minimum), the authors wish to find the input $\mathbf{x}_0$ where $\mathrm{E}[QL(\mathbf{x}_0 + \mathbf{z})]$ is minimal (the best operating minimum). Since $QL$ is a polynomial of degree 4 in $\mathbf{x}$, we can write this as

$$\begin{aligned} \mathrm{E}[QL(\mathbf{x}_0 + \mathbf{z})] &= \mathrm{E}[QL(\mathbf{x}_0) + \mathbf{z}^T \nabla QL(\mathbf{x}_0) \\ &\quad + \frac{1}{2}\mathbf{z}^T \nabla^2 QL(\mathbf{x}_0)\mathbf{z} \\ &\quad + \text{3rd and 4th order terms}]. \end{aligned} \tag{6}$$

where $\nabla$ denotes gradient and $\nabla^2 QL(\mathbf{x}_0)$ is the Hessian matrix of $QL$ at $\mathbf{x}_0$.

The following assumptions are made concerning the errors $\mathbf{z}$ in $\mathbf{x}_0$:

- $\mathbf{z}$ is independent of $\mathbf{x}_0$
- All $z_i$ are independent.
- $\mathrm{E}[z_i] = 0$.
- $\mathrm{E}[z_i^3] = 0$.

With these assumptions and the fact that $\mathbf{x}_0$ is a local minimum, Equation (6) reduces to

$$\mathrm{E}[QL(\mathbf{x}_0 + \mathbf{z})] = QL(\mathbf{x}_0) + \frac{1}{2}\sum_{i=1}^{k} \frac{\partial^2 QL(\mathbf{x}_0)}{\partial x_i^2} \mathrm{E}[z_i^2] + C. \tag{7}$$

The first order term is zero because $\mathrm{E}[z] = 0$ and because $\nabla QL(\mathbf{x}_0) = 0$ since $\mathbf{x}_0$ is a local minimum. The third order term is zero since $\mathrm{E}[z_i^3] = 0$ and the fourth order term is a constant; it does not need to be considered in the minimization.

The best operating minimum, chosen from the local minima $\mathbf{x}_0$, is the minimum where

$$\mathrm{E}[QL(\mathbf{x}_0 + \mathbf{z})] = QL(\mathbf{x}_0) + \frac{1}{2}\sum_{i=1}^{k} \frac{\partial^2 QL(\mathbf{x}_0)}{\partial x_i^2} \mathrm{E}[z_i^2] \tag{8}$$

is minimal.

## 1.2 Analysis of the Model

### 1.2.1 Analysis of the Regression Analysis Part

The regression analysis part of the paper, as described in the previous section, is based upon two main assumptions.

The first one is that a quadratic function is a "good guess" for the actual response function, describing how the output variables $y_i$ depend upon the input variables $x_i$. No reasons are given why this would actually be the case, and no numerical data are provided for to check this statement.

The second underlying assumption is less obvious. Notice that the coefficients $\beta_i$ and $\beta_{ij}$ of the quadratic regression function are determined by minimizing the sum of the squared lengths of *vertical* deviations (if $k = 1$). This gives the impression that the input variables $x_i$ are known exactly !
However, consider a practical situation in which a scientist wants to carry out an experiment. He may wish to add an amount $x_i$ of chemical type $i$ to a solution – but as a consequence of measurement errors he is bound to add an amount $x_i + \delta_i$, where $\delta_i$ may be positive or negative. Similarly, the measured output $y_i'$ equals the real output $y_i$ plus a measurement error, say $\epsilon_i$.
This means that one actually wants to find a regression function satisfying

$$\mathbf{y}' = f(\mathbf{x}') + \mathbf{e},$$

where $\mathbf{x}' = \mathbf{x} + \delta$, $\mathbf{y}' = \mathbf{y} + \epsilon$, $\delta$ and $\epsilon$ denoting the vector of errors in the input variables and in the output variables respectively.
Intuitively spoken, taking $k = 1$ again, on does not wish to minimize the sum of the squared lengths of *vertical* deviations, but something like the sum of the squared lengths of *orthogonal* deviations, as depicted in Figure 2 ! However, more thorough mathematical investigations seem to be appropriate here.

### 1.2.2 Analysis of the Minimization Part

Many assumptions are made in the minimization part of the paper. We will list and comment them all.

The most surprising one is that the authors all of a sudden seem to forget that the values of the coefficients $\beta_i$ and $\beta_{ij}$ appearing in the response function are actually stochastic. They consider them to be known exactly. As a consequence, the possible "errors" in these coefficients do not play any role in the sequel, which might not be realistic. Such "errors" might occur if the samples of input parameters $x_i$ used to determine the $\beta_i$ were not representative for the actual values of the input parameters occurring in practice.

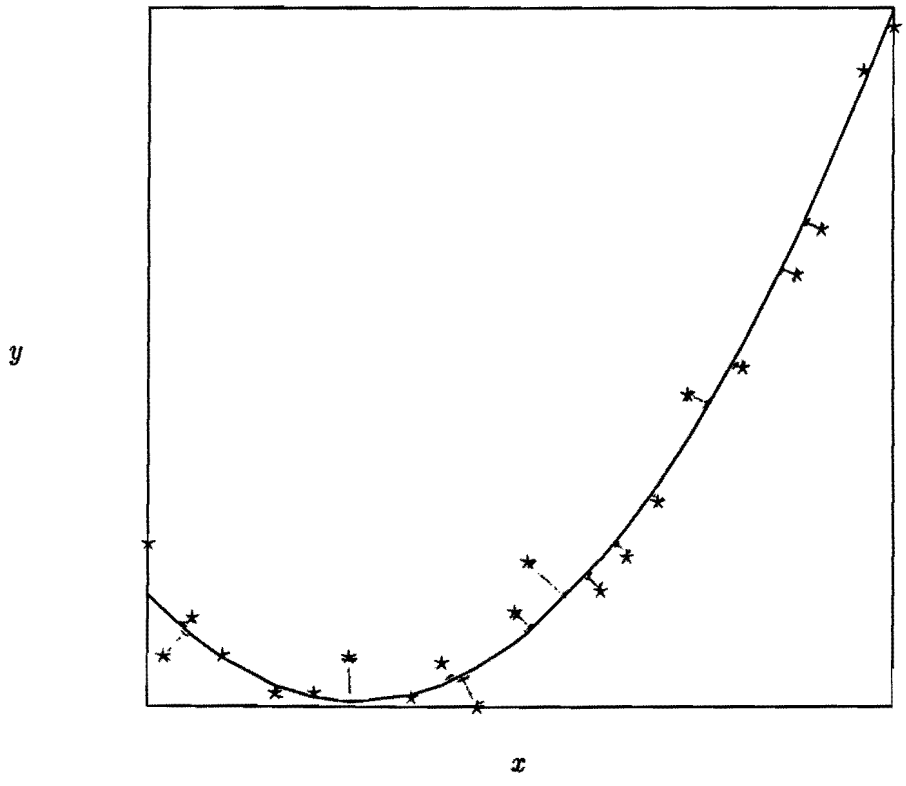



Figure 2: Variant of quadratic regression ($k = 1$). The sum of the squared lengths of the *orthogonal* deviations is minimized.

On the other hand, the input variables $\mathbf{x}$ are now supposed to be disturbed with errors $\mathbf{z}$, the distribution function of which is assumed to be known. The $z_i$ are assumed to be independent, and their first and third moments are set to zero. Furthermore, the authors propose to let $\mathbf{z}$ be independent of $\mathbf{x}$.

Thus, the problem considered in the second part of the paper does not seem to correspond completely with the one suggested in the first part. Instead of minimizing the quality-loss function assuming the $\beta$-coefficients to be stochastic, this minimization is performed under the assumption that they are exactly known, while the input variables are disturbed.

Furthermore, the assumptions concerning the errors $\mathbf{z}$ in the input variables may be commented.
Although it seems reasonable to assume the mean of the $z_i$ to be zero, why is their third moment ? The independence of the $z_i$ has to do with the way in which the input parameters $\mathbf{x}$ are controlled. If, for instance, one lever controls the input amount of six $x_i$-values, these are certainly not independent. Finally, one may wonder whether it is reasonable to assume $\mathbf{z}$ to be independent of $\mathbf{x}$. Another – reasonable – assumption would seem to be that the errors $\mathbf{z}$ are larger of the input values $\mathbf{x}$ increase. As an example, consider the task of filling a glass or a bucket with water, where the water is the only input variable $x$. Filling the glass, one is likely to spill a few drops, while filling the bucket, it is quite well possible that one spills a few glasses. Thus, the spilling error $z$ is not independent of the input value $x$ !

Another assumption which underlies the reasoning in the minimization part of the paper is that the optimal output values $t_r$ are known. This means that, when using the minimization method proposed in the paper, one must be able to quantify exactly what the desired output is. Notice that this is not likely to be the case in the example of developing pictures: one customer may wish his pictures to have very bright colors, paying less attention to the contrast, while another may be specially interested in the contrast. Thus, it is questionable whether the chosen quality-loss function, and the parameters in it, are realistic.

In order to determine the expected value of the quality-loss function in case the input vector $\mathbf{x}$ is stochastic, the authors of the paper have made a Taylor-expansion of $QL(\mathbf{x}_0+\mathbf{z})$ around $\mathbf{x}_0$. We asked ourselves whether making this kind of expansions is allowed when the variables involved are stochastic.

Finally, one may wonder whether the method used to find the best operating minimum is good. The idea is clear: as it is not exactly known how large the input values $\mathbf{x}$ are, one can neither predict the magnitude of the output values $\mathbf{y}$, nor the corresponding value of the quality-loss function – and so the authors propose to minimize its expectation.
But what exactly do they define to be the expectation ? As Figure 3 shows, it does

not seem to be appropriate to take

$$\mathbf{E}[QL(\mathbf{x}_0 + \mathbf{z})] = \int_{-\infty}^{\infty} QL(\mathbf{x}_0 + \mathbf{z}) dF_z(\mathbf{z}). \tag{9}$$

where $F_z(\mathbf{z})$ is the probability distribution function of $\mathbf{z}$ In fact, for some possible quality-loss functions, this integral may be infinite, independent of the chosen local minimum $\mathbf{x}_0$. If the quality-loss function is a fourth degree polynomial in $\mathbf{x}+\mathbf{z}$, and the density function shows the behaviour of $1/|\mathbf{z}|^2$ for $|\mathbf{z}| \to \infty$, the integral in (9) will diverge.

On the other hand, one may argue that in practical situations the errors $\mathbf{z}$ are normally distributed with mean zero. In that case, whatever polynomial quality-loss function one uses, the expectation $\mathbf{E}[QL(\mathbf{x}_0 + \mathbf{z})]$ is finite.

This suggests that it would be wise to assume the distribution functions of the $z_i$ to have a compact support, i.e. to assume these functions to be non-zero on a finite interval of the real axis only.

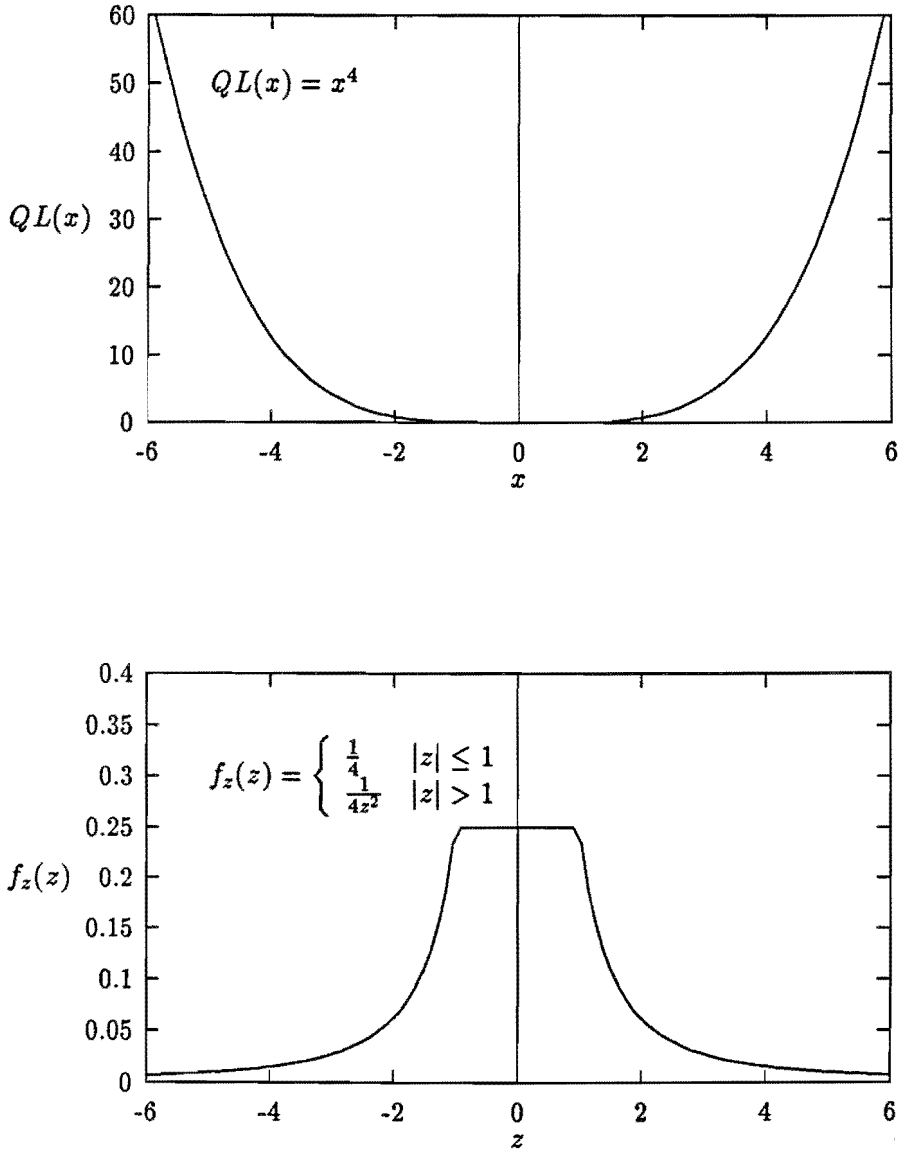



Figure 3: Using the "usual" definition of expectation does not seem to be appropriate when certain quality-loss functions may occur.

# 2 Looking for Possible Mistakes

As explained in the previous section, quite a lot of – sometimes not explicitly mentioned – assumptions underlay the paper "Experimental Design and Quality-Loss Function". Some of these assumptions do seem to be quite questionable indeed. In the sequel we will elaborate on them, and possible provide for alternative assumptions.
Unfortunately, though, it will not be possible to give any qualitative comparison between the suggested alternatives and the methods the authors use: we do not have any "real-life" data.

## 2.1 Possible Mistakes in the Regression Analysis Part

As mentioned before, in the regression analysis part of the paper "Experimental Design and Quality-Loss Function", two assumptions are made which seem to be dangerous. First, quadratic regression may not provide for a good response function; second, as the input variables $x_i$ are not exactly known upon determining this response function, one could expect the usual least square method used in regression analysis not to yield the proper results.

As we have not been provided with any numerical data, we can not check whether the quadratic response function is good or not. Still, it seems to be quite a coincidence that *all* the output variables $y_i$ in photographic practice depend in a quadratic way on *all* the input variables $x_i$.

As far as the least square method is concerned, we can propose an alternative.
Suppose a (quadratic) response function describing the relation between input and output variables is given – as in Figure 4.
For the sake of simplicity, let us consider the case where $k = 1$. According to the above response function, an input $x_0$ corresponds to an output $y_0$, and all other possible inputs $x$ yield a larger output $y$.
Now suppose one tries to input $x_0$. In practice, the actual input amount will appear to be $x + \delta$, $\delta$ denoting a measurement error. As a consequence, the resulting output will not be $y_0$, as desired, but some $y + \epsilon = f(x + \delta)$. If one repeatedly tries to input $x_0$, the corresponding output values will all lie *above* the graph of the response function, as shown in Figure 4. Furthermore, there will be more input-output pairs of which the graphical representation lies close to the response function. The "cloud" of input-output pairs can be expected be egg-shaped, is displayed in the figure.
What happens if one wants to study the effect of using an input amount $x^*$ ? The real input to the system will be $x^* + \delta$, where $\delta$ may be positive or negative – and

the corresponding output is $y^* + \epsilon = f(x^* + \delta)$, where $\epsilon$ and $\delta$ can be seen to have the same sign ! Thus, repeatedly trying to input $x^*$ now results in a "cloud" of output values lying *above and under* the response function graph. The cloud will look more like an ellipse than like an egg; we again refer to Figure 4.

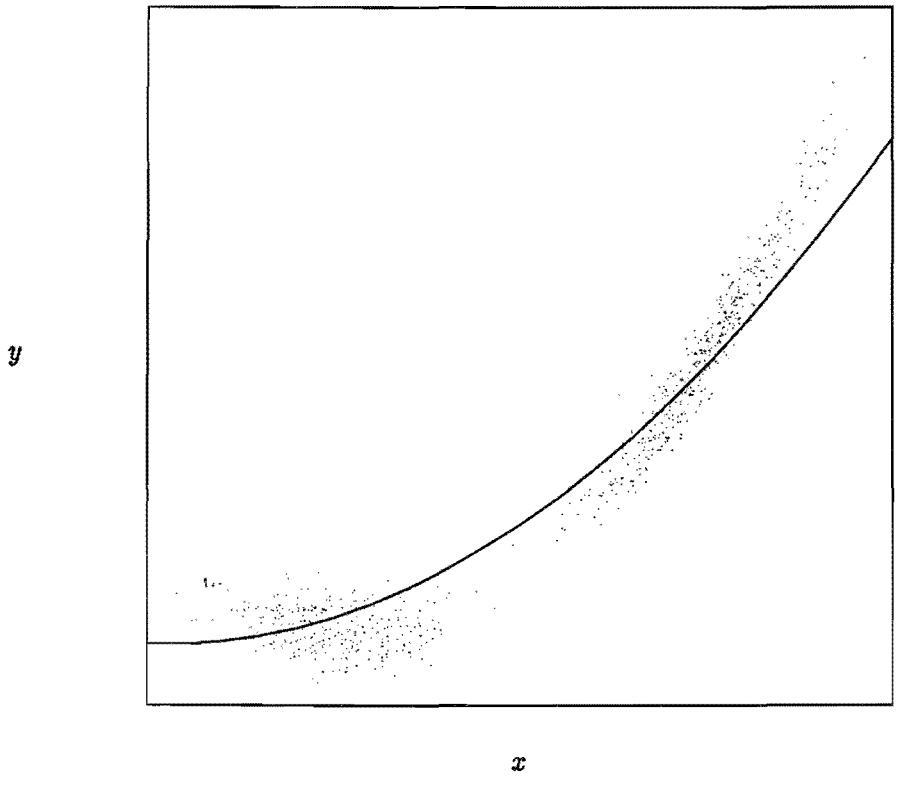



Figure 4: The response function and "clouds" of input-output pairs.

In reality, the experiments are carried out in order to find the response function, which is therefore not known a priori. The practical situation is therefore the one depicted in Figure 5: after a number experiments, a number of "clouds" has been found showing how the output of the system depends on its input.
The graph in Figure 5 roughly indicates the result of a least square fitting: it passes as close as possible through the the middles[1] of the input-output "clouds".
Comparing Figures 4 and 5, however, it appears that this is not the fitting we want ! The response function should pass through the middle of a "cloud" if it is ellipse-

[1] If each point in a "cloud" is assumed to have mass one, the middle of the "cloud" corresponds to its centre of gravity

shaped, and underneath if the "cloud" looks like an egg.

We therefore propose the following alternative to the least square regression method. Given the "clouds" of input-output pairs, determine their shape by considering the distances between the corresponding points. Introduce a coefficient indicating whether the "cloud" is ellipse - or egg-shaped. This coefficient will serve as a correction factor, moving the actual middle of the "cloud" downward the more it is egg-shaped.

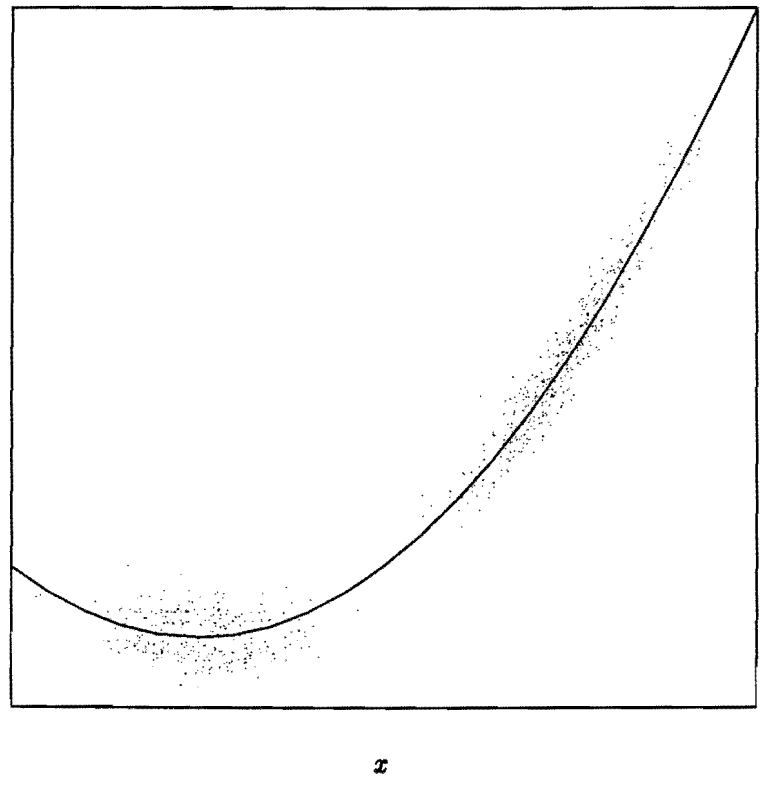



Figure 5: Alternative regression method for determining the response function.

## 2.2 Possible Mistakes in the Minimization Part

From the section Analysis of the Minimization Part, a number of possible mistakes can easily be pointed out:

1 The $\beta$-coefficients of the (quadratic) response function are in practice not exactly known, but stochastic. This may influence the results.

2 The errors z in the input x may depend on the input. Furthermore, the $z_i$ could be interdependent, and their first and third moments could differ from zero in practice.

3 It may happen that the optimal output values $t_r$ are not exactly known. In the photographic example, the $t_r$ are bound to have been determined visually, which can not be expected to yield one theoretical optimum, but rather a range within the optimum must lie.

4 Is it allowed to make a Taylor-expansion of a function the variables of which are stochastic ?

5 Using the usual definition of expectation does not seem to be appropriate when certain quality-loss functions may occur, as shown in Figure 3.

As far as point 5 is concerned, we want to make sure that the integral in (9) does not diverge. If the quality-loss function is polynomial, we saw that assuming a normal distribution for z solves the problem. In general, any density function which decreases exponentially for $|\mathbf{z}| \to \infty$ satisfies.
If the quality-loss function is not polynomial, however, more restrictions need to be made concerning the choice of a density function. Allowing all possible continuous quality-loss functions, a possible satisfactory solution to the problem of infinite expectation would be to assume the density function of z to have a compact support. An example is given in Figure 6.
In the sequel we assume the density function of z to be such that the expectation (9) is finite. Furthermore (considering point 2) it seems to be reasonable to assume its first moment to be zero. If the expectation of the $z_i$ is not zero, a structural mistake is made when measuring the amounts of input $x_i$ – as may be the case when a machine does not work properly.

We will consider the remaining possible mistakes listed in the beginning of this section simultaneously.

To do so, we notice that it seems to be quite superfluous to make a Taylor-expansion of a function which is known to be a fourth degree polynomial. In fact, the expansion will look exactly the same as the function itself !
From (4) and (5) we have

$$\begin{aligned} QL(\mathbf{x}) &= L(\mathbf{y}(\mathbf{x})) \\ &= \sum_{r=1}^{n} \{(y_r(\mathbf{x}) - t_r) \cdot w_r\}^2 \end{aligned}$$

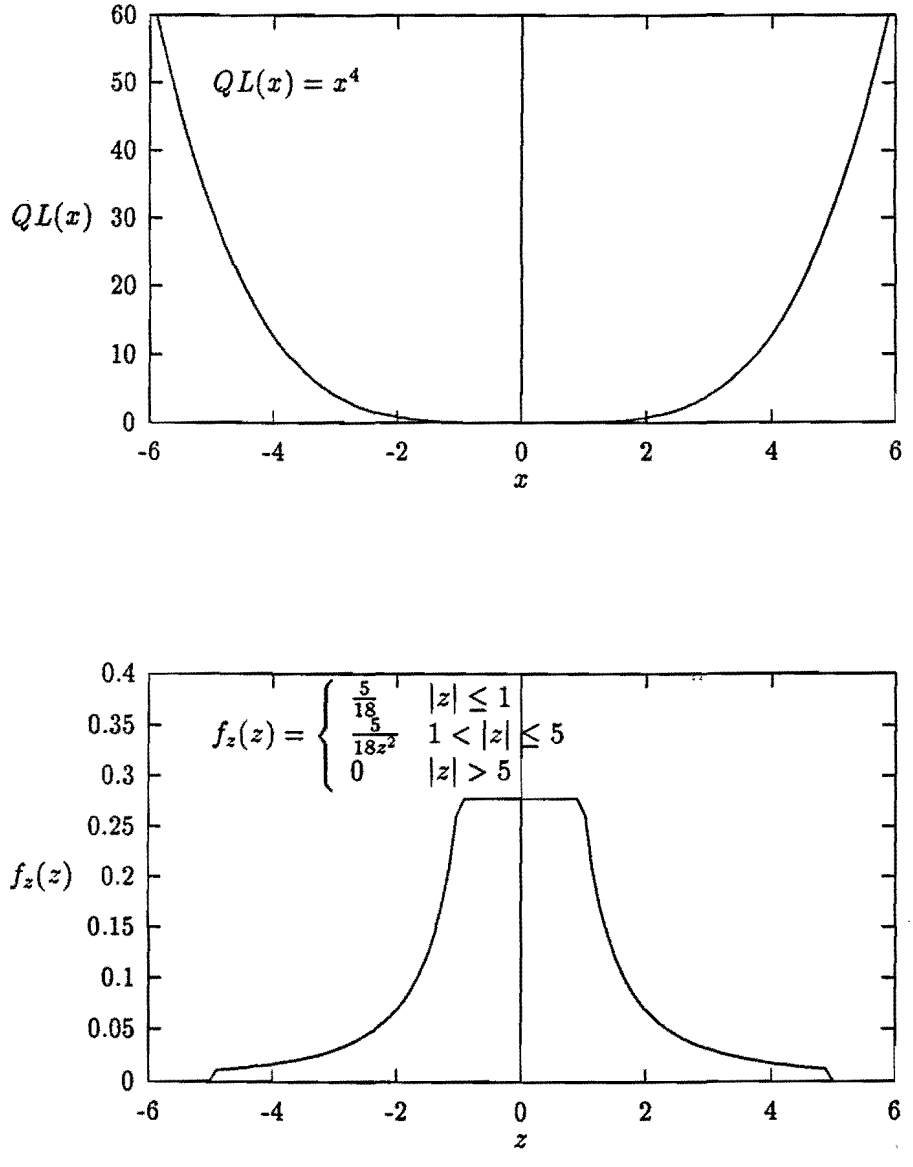



Figure 6: An "ugly" quality-loss function and a density function with compact support for which the expectation (9) is finite.

$$= \sum_{r=1}^{n}\{(\beta_{0r} + \sum_{i=1}^{k}\beta_{ir}x_i + \sum_{i,j=1(i\leq j)}^{k}\beta_{ijr}x_ix_j - t_r)\cdot w_r\}^2. \qquad (10)$$

Setting $\mathbf{x}$ to $\mathbf{x}_0 + \mathbf{z}$ as in (6), we find (denoting the entries of $\mathbf{x}_0$ as $x_1, x_2, \ldots, x_k$.

$$\begin{aligned}
& QL(\mathbf{x}) \\
= & \sum_{r=1}^{n}\left\{\left(\beta_{0r} + \sum_{i=1}^{k}\beta_{ir}(x_i+z_i) + \sum_{i,j=1(i\leq j)}^{k}\beta_{ijr}(x_i+z_i)(x_j+z_j) - t_r\right)\cdot w_r\right\}^2. \\
= & \sum_{r=1}^{n} w_r^2\cdot\{\beta_{0r}^2 + (\sum_{i=1}^{k}\beta_{ir}(x_i+z_i))^2 + (\sum_{i,j=1(i\leq j)}^{k}\beta_{ijr}(x_i+z_i)(x_j+z_j))^2 + t_r^2 \\
& +2\beta_{0r}\sum_{i=1}^{k}\beta_{ir}(x_i+z_i) + 2\beta_{0r}\sum_{i,j=1(i\leq j)}^{k}\beta_{ijr}(x_i+z_i)(x_j+z_j) - 2\beta_{0r}t_r \\
& +2(\sum_{i=1}^{k}\beta_{ir}(x_i+z_i))(\sum_{i,j=1(i\leq j)}^{k}\beta_{ijr}(x_i+z_i)(x_j+z_j)) - 2t_r\sum_{i=1}^{k}\beta_{ir}(x_i+z_i) \\
& -2t_r\sum_{i,j=i(i\leq j)}^{k}\beta_{ijr}(x_i+z_i)(x_j+z_j)\} \\
= & \sum_{r=1}^{n} w_r^2\cdot\{\beta_{0r}^2 + \sum_{i,j=1}^{k}\beta_{ir}\beta_{jr}(x_i+z_i)(x_j+z_j) \\
& + \sum_{i,j=1(i\leq j)}^{k}\sum_{s,t=1(s\leq t)}^{k}\beta_{ijr}\beta_{str}(x_i+z_i)(x_j+z_j)(x_s+z_s)(x_t+z_t) + t_r^2 \\
& +2\beta_{0r}\sum_{i=1}^{k}\beta_{ir}(x_i+z_i) + 2\beta_{0r}\sum_{i,j=1(i\leq j)}^{k}\beta_{ijr}(x_i+z_i)(x_j+z_j) - 2\beta_{or}t_r \\
& +2\sum_{i,j=1(i\leq j)}^{k}\sum_{s=1}^{k}\beta_{ijr}\beta_{sr}(x_i+z_i)(x_j+z_j)(x_s+z_s) - 2t_r\sum_{i=1}^{k}\beta_{ir}(x_i+z_i) \\
& -2t_r\sum_{i,j=1(i\leq j)}^{k}\beta_{ijr}(x_i+z_i)(x_j+z_j)\}. \qquad (11)
\end{aligned}$$

What can we see from this – quite impressive – formula ?
First, it allows us to check whether or not it is justified to make a Taylor-expansion of the quality-loss function around $\mathbf{x}_0$. Taking the expectation of (11) and simplifying by means of the assumptions listed on p.7 should yield the same result as displayed

in (7). Second, using formula (11), we can try to find out what happens if the mentioned assumptions are not made.

We will not go through the complete reduction of (11) to (7) in case the assumptions are made. Notice, however, that the term $QL(\mathbf{x}_0)$ appears easily from (11) by considering only those terms (and factors) in which no components of the error vector $\mathbf{z}$ appear. The terms containing exactly one or three components of $\mathbf{z}$ vanish as a consequence of the assumptions made. Can you imagine how happy we were when we found out that the terms in which two $\mathbf{z}$-components appear do indeed add up to

$$\frac{1}{2}\sum_{i=1}^{k}\frac{\partial^2 QL(\mathbf{x}_0)}{\partial x_i^2}\mathrm{E}[z_i^2] \ ?$$

As a side-result of this glorious calculation we find

$$C = \sum_{r=1}^{n} w_r^2 \sum_{i=1}^{k} \mathrm{E}[z_i^4], \tag{12}$$

which is indeed a constant.
The above results show that it is indeed justified to use a Taylor-expansion of the quality-loss function, even if the variables involved are stochastic.

How can we determine the expected value of (11) in case we do not make the assumptions listed on p.7 ? And what happens if we consider the $\beta$-coefficients to be stochastic, and if we take into account the possibility that the optimal values $t_r$ may be subject to stochastical fluctuations ?
We have to confess we did not find any satisfactory answer to this question. "Just" taking the expectation – assuming the density functions of all stochastical variables to be known – is practically impossible. At first sight it would seem reasonable to neglect terms in which two or more $\mathbf{z}$-components occur, as the $z_i$ can be assumed to be small compared to the $x_i$. Unfortunately, however, as $\mathbf{z}$ and $\mathbf{x}$ are dependent, each particular $z_i$ can only supposed to be small in comparison with the corresponding $x_i$ – and not to the other input parameters !

Thus, although the minimization part of the paper "Experimental Design and Quality-Loss Function" does not seem to be quite satisfying from a mathematical point of view, we can not provide for any suitable alternatives without knowing more about "real-life" data.

# 3 Conclusions

In this report, we considered the paper "Experimental Design and Quality-Loss Function" about minimizing a quality-loss function the value of which depends on a set of stochastical input variables.
The authors of the paper propose to use regression analysis to find out how exactly the quality-loss function depends on the input variables. This dependence is expressed in terms of a response function.
Having found the response function, a minimization method is described to find the input vector $\mathbf{x}$ such that the expected value of the quality-loss function is as small as possible.

We analysed both the regression analysis and the minimization part of the paper, finding quite a lot of questionable assumptions.

As far as the regression analysis part is concerned, we wondered whether the least square method was suited for the problem under consideration, and, if so, whether or not quadratic regression was appropriate.
We presented a possible alternative for the traditional least square method, which seems to more adapted to the problem. It was not possible to study the quality of quadratic regression in detail, because we did not have any numerical data.

In the minimization part of the paper, the authors did not take into account that a number of variables used actually originated from regression analysis. Instead of treating these variables as stochastic variables, they were considered to be known constants. The authors also made quite heavy assumptions about the stochastical behaviour of the input variables. Finally, some of the mathematical methods and definitions used were questionable – at least to us.
We checked whether or not the mathematics were correct in the cases where questions raised, and we introduced an alternative definition of expectation. We were not able to find out if the minimization method still would yield satisfactory results if the assumptions on the stochastical behaviour of the input variables were dropped, or if the coefficients resulting from the regression analysis were indeed considered to be stochastical.

We wish to emphasise that the lack of actual data, originating from a "real-life" problem, made the "inverse modelling" procedure rather difficult. The modelling itself – by the authors of the paper – could not be studied, and only little insight could be gained into the influence of the stochastical aspect of the input variables.

# A MODEL FOR CRYSTAL PRECIPITATION

# Contents

# Preface

This report is written for the Mathematical Modelling course given at the Technical University of Eindhoven for the students of the postdoc-education program "Mathematics for Industry". For this course we studied the article "Conservation Laws in Crystal Precipitation", a joint work by P.E. Castro, A.E. Cha-Lin, D.S. Ross and P.H. Karpinski [Cas87]. In this report we present a model which can be found in this article. Furthermore we present some numerics that we did to check some theoretical results.

# 1 Introduction

Many industrial processes depend upon particular components for their effectiveness. In the photographic business for instance think of particles, so called "silver halide grains", where precipitation of these particles in photographic emulsions is the key to high photographic imaging. Two things are important in this case: the distribution of these particles and their morphologies. These particles appear as crystals in the emulsions where they tend to grow. Therefore it is important to describe the evolution process of crystals by mathematical models. By doing so, one will gain insight in the influence of quantities such as the starting concentration and injection of new crystals.

Of what sort of crystals do we have to think? We think of crystals which are three dimensional. For instance cubes, balls or polygons with a certain thickness. So we try to describe the evolution process of such crystals in a situation where solved matter will crystallize when the concentration reaches the saturation concentration, and crystals solve when the concentration decreases. A simple example is the precipitation of salt in water.

# 2 The general model

In our observation we are interested in the number of crystals in a special form at time t. Therefore we describe the crystals by location and a few physical characteristics. So the spatial vector is extended with m physical parameters:

$$R = (r_1, r_2, .., r_{m+3})$$

Now $n(R,t)$ is the density of crystals with parameter $R$, the amount on location $(r_1, r_2, r_3)$ with characteristics $(r_4, .., r_{m+3})$ expressed per unit mass of solvent. Characteristics can be features as diameter and thickness.

To be more specific, when we introduce $dR$, we are able to describe the number of crystals in a volume. The meaning of $dR$ is understood by comparing $dR$ with the notion of $dx$ in $\mathbb{R}^3$. When $m = 0$ we obtain $dR = dx$, this is a volumeblock with coordinates $x_i$ in the interval $[x_i, x_i + dx_i]$ $(i = 1,2,3)$. So $dR$ is a sort of block with the spatial variables in a volumeblock and the characteristics in the intervals $[r_i, r_i + dr_i]$. Now $n(R,t)dR$ is the number of crystals in $dR$ at time $t$ per unit mass of solvent. Hence $n(R,t)$ is expressed per unit mass of solvent and per unit of length to the power $m + 3$.

The derivative of $R$, $v = dR/dt$, is the phase velocity of crystals, this means the rate of change both in space and in characteristics.

For describing the crystal density $n$, the density can be compared with fluid density. Thus the continuity equation for crystals can be formed, whereby $B$ and $D$ are introduced to denote the birth and death of crystals.

$$\frac{\partial n}{\partial t} + div(nv) = B - D \tag{1}$$

Birth of crystals means the precipitation of solved matter into new crystals. Analogous to birth, death stands for the solving of existing crystals. Roughly said the continuity equation claims that the change in density in phase $R$ on time $t$ plus the change in crystals due to changes in form (as well in location as in characteristics) is equal to the birth minus death of crystals.

For an understanding of the whole volume of fluid containing solved matter and crystals, it is useful to make some assumptions. First we assume an uniform distributed precipitation, which implies $n(R,t)$ is constant over $r_1..r_3$. To have some influence on the process of crystallization we introduce $F$ for the rate of injection of crystals. Then we can form a continuity equation for the entire volume of fluid. Because $n(R,t)$ is defined as the number of crystals per unit

mass of solvent the continuity equation holds for $n(R,t)S(t)$, where $S(t)$ is the total mass of solvent in the crystal. Furthermore $v = (v_x, v_i)$, with $v_x$ denoting the velocity in space and $v_i$ denoting the velocity in characteristics. Then $div(nv) = div(nv_x) + div(nv_i) = -F.S(t) + div(nv_i)$ and for the entire volume (1) results in

$$\frac{\partial n}{\partial t} + div(nv_i) + \frac{n}{S}\frac{dS}{dt} = B - D + F \qquad (2)$$

In this formula $v_i$ will depend on the concentration of solved matter. The rates $B$, $D$ and $S$ are endogenous, $F$ is to control. With a specification of these variables $n$ can be determined by (2).

# 3 Ostwald ripening

In order to observe the crystallization process, one has to specify five variables, according to the previous paragraph. Each specification of these variables corresponds with a special sort of precipitation of crystals. In this section we will simplify the problem described in (2) by taking easy functions for the five variables. The resulting problem corresponds with the so called process of Ostwald ripening, and is described by N.S. Tavare [Tav87].

The first assumption is the absence of birth and death of crystals, and the absence of injection. So once a crystal, always a crystal, only the characteristics are varying. Thus $B = D = F = 0$. Furthermore it is supposed that the amount of solvent in the crystals is constant: $dS/dt = 0$.

For the characteristics of the crystal we only observe the radius L. Now the parameter vector becomes $R = (r_1, r_2, r_3, L)$. So $v_i$ is written as $dL/dt$ and is a function $G$ depending on $L$ and $t$.

$$\frac{dL}{dt} = G(L,t)$$

For $n$ being constant over space as in (2) the density is simplified to $n(R,t)dR = n(L,t)dL$, where $n(L,t)dL$ is the number of crystals per unit mass of solvent with the radius in $[L, L+dL]$ at time t.

With the above assumptions the continuity equation is

$$\frac{\partial n}{\partial t} + \frac{\partial (nG)}{\partial L} = 0 \tag{3}$$

Function $G$ describes the growth (positive and negative) of the crystals. In the Ostwald ripening small crystals tend to dissolve, and large crystals tend to grow. This means on time $t$ there is a critical $L^*$ with $G(L^*,t) = 0$. Further $G < 0$ for $L(t) < L^*(t)$ and $G > 0$ for $L(t) > L^*(t)$.

Such a function is found by relating the growth to concentration of crystals and a function depending on $L$ which is equal to the concentration for $L = L^*$.

$$G(L,t) = f(\, c(t) - g(L)\,), \quad g(L^*) = c(t)$$

Where $f$ is a certain function. Now $g(L)$ is given by the Gibbs-Thomson relation:

$$g(L) = c^* e^{\Gamma_D/L} \tag{4}$$

In (4) $\Gamma_D$ is a physical constant: $\Gamma_D = 4\sigma v/RT$ containing the surface energy, the molecular volume, gas constant respectively the temperature. $L^*$ is to determine by $g(L^*) = c(t)$. We find

$$L^*(t) = \frac{\Gamma_D}{\ln \frac{c(t)}{c^*}}$$

Next step in formulating the problem is to determine the relation between $G$ and $c(t) - g(L)$; to find an expression for $f$. In [Cas87] and [Tav87] conventional power law growth kinetics are proposed, and coefficients are empirically derived. Resulting in

$$G(L,t) = \begin{cases} k_g(\, c(t) - c^* e^{\Gamma_D/L}\,)^g & \text{for } L \geq L^*(t) \\ -k_d(\, c^* e^{\Gamma_D/L} - c(t)\,)^d & \text{for } L < L^*(t) \end{cases} \tag{5}$$

$$1 \leq g \leq 2,\ 1 \leq d \leq 2,\ k_g, k_d \geq 0$$

A problem in this specification lies in the determination of $c(t)$. For the concentration is depending on $n(L,t)$:

$$c(t) = c_0 + \rho k_v \left( \int_0^\infty L^3 n(L,0) dL - \int_0^\infty L^3 n(L,t) dL \right) \tag{6}$$

The constants $\rho$ and $k_v$ are the mass density respectively the ratio of crystal volume (for a crystal of size $L$) and $L^3$.

Setting out the above specifications of the crystalization we obtain a first order partial differential equation for $n(L,t)$ with starting condition $n(L,0)$:

$$\frac{\partial n}{\partial t} + G\frac{\partial n}{\partial L} = -n\frac{\partial G}{\partial L} \tag{7}$$

with starting condition $n(L,0)$:

$$n(L,0) = n_0(L) \geq 0$$

The complication in $c(t)$ arises because $n$ is depending on $G$ by (7), $G$ depends on $c$ by (5), and back to the starting point, $c$ is related to $n$ by (6).

# 4 Numerical methods

Because the continuity equation (7) in Tavare's model cannot be solved analytically we have tried to solve it by numerical methods. Two finite difference methods have been tried, an explicit method and an implicit method and a thirth method based on characteristics. In the following the results of all three methods will be shown, together with an idea for a reparametrisation.

## 4.1 Finite differences, explicit method

The first finit difference method is an explicit method. This method is not stable, which can easily be shown. We used the following inverse T-shape discretisation for the derivatives in grid point $(i, j)$.

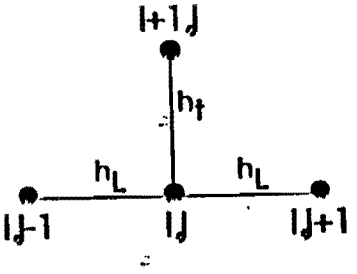



Figure 1: Inverse T-shape discretisation molecule

$$n_t = \frac{n_{i+1,j} - n_{i,j}}{h_t}$$

$$n_L = \frac{n_{i,j+1} - n_{i,j-1}}{2h_L}$$

where $i$ is the i'th grid point in the t-direction and $j$ is th j-th grid point in the $L$-direction.

The discretised continuity equation yields:

$$\frac{n_{i+1,j} - n_{i,j}}{h_t} + G\frac{n_{i,j+1} - n_{i,j-1}}{2h_L} + n_{i,j}G_L = 0 \qquad (8)$$

To check the validity of this discretisation scheme the simple PDE

$$u_t + u_L = 0$$
$$u(L,0) = sin(2\pi L), \quad L < 1$$
$$u(L,0) = 0, \quad L > 1$$

is implemented in Matlab. The result is shown in figure (2). It was expected that the sine-wave would be preserved and move to the right in time. But using the discretisation described above the wave is damped after only a few time-steps. This phenomena is called *numerical diffusion.* A close observation also shows *numerical dissipation*, the phenomena of very little waves on both sides of the major curve.

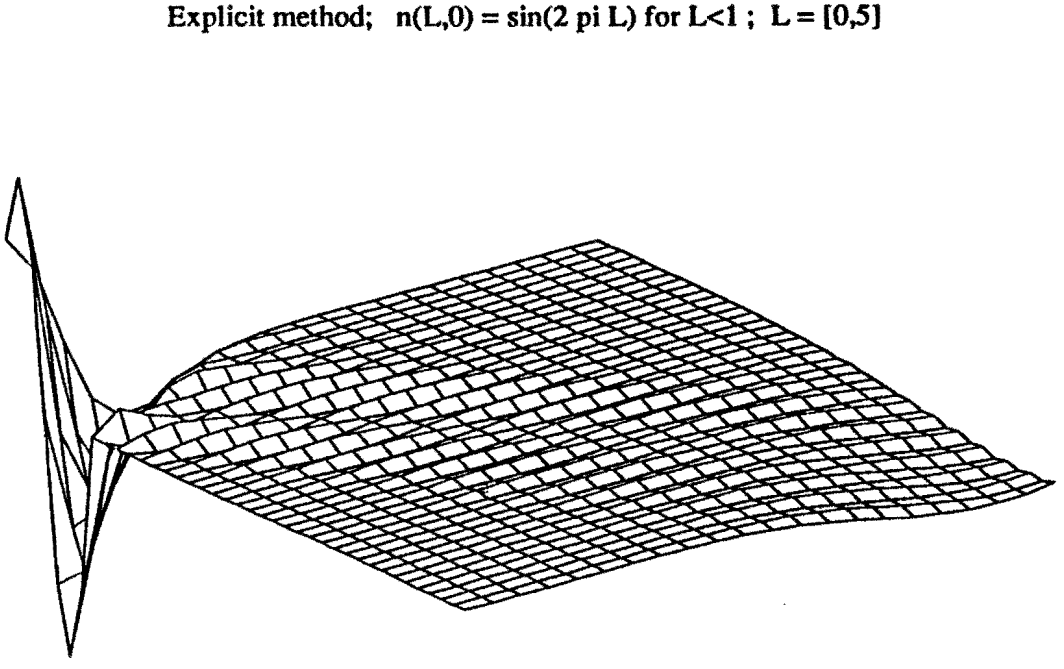



Figure 2: The test problem for the explicit discretisation scheme

If we take for the constants in Tavare's model the following values:

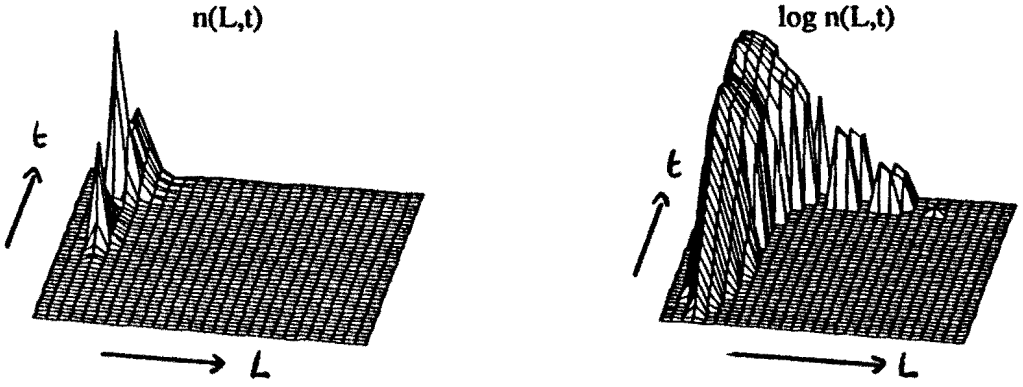
$$
\begin{aligned}
\Gamma_D &= 8 * 10^{-5} \\
g &= 1.5 \\
d &= 1.5 \\
k_g &= 7.9 * 10^{-8} \\
k_d &= 7.9 * 10^{-8} \\
c^* &= 0.1 \\
\rho &= 2000
\end{aligned}
$$

$$k_v = 0.52$$

and an uniform starting distribution:

$$u(L,0) = 0.5, \forall L$$

we get the following result (see figure 3)




tmax = 2 dt = 0.05
Lmax = 0.0001 dL = 3.333e-006
Lmin = 0
L0 = 4.971e-005
n0 = 0.5
#L = 30
#t = 40

Figure 3: Solution of Tavare's problem using an explicit discretisation scheme

This result is not accurate, as we will see later. A large peak appears for small L. The peak value is of order $10^6$. It can be proven quite easily that every value n(L,t) can never be bigger than the maximum of n(.,0).

## 4.2 Finite differences, implicit method

As a second attempt an implicit discretisation scheme is tried. The following T-shape discretisation molecule is used for the derivatives in grid point $(i, j)$

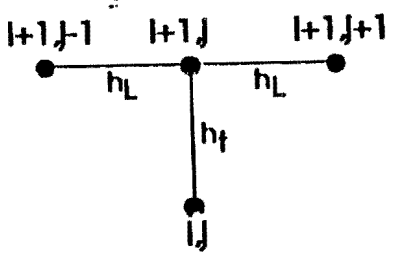



Figure 4: T-shape discretisation molecule

$$n_t = \frac{n_{i+1,j} - n_{i,j}}{h_t}$$

$$n_L = \frac{n_{i+1,j+1} - n_{i+1,j-1}}{2h_L}$$

For the test problem also numerical diffusion is present, but it has less influence than for the explicit schemes (see figure (5)).

For Tavare's model the following result is obtained. (figure (6))

We see that all values of n(L,t) are equal or less than the values at $t = 0$. This will coincide with the results in the section about the characteristics.

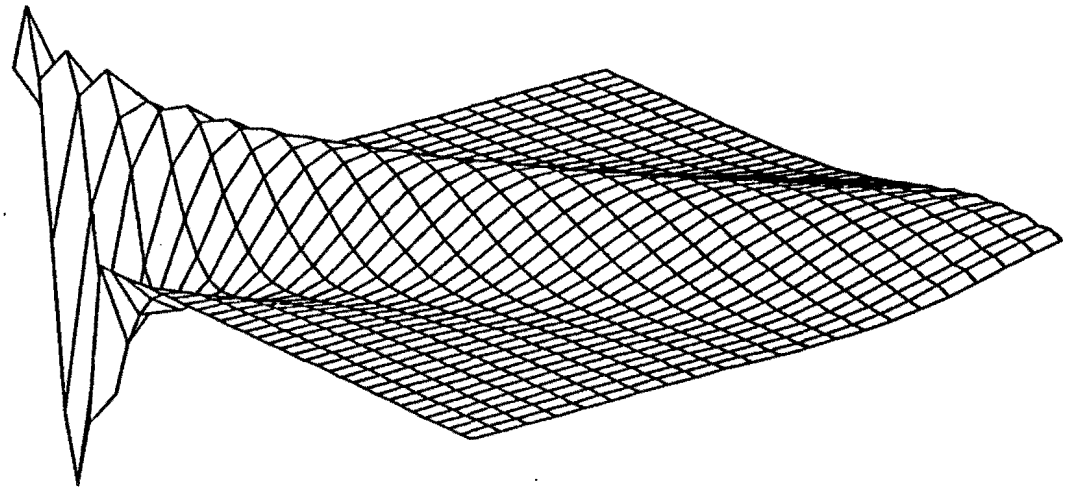
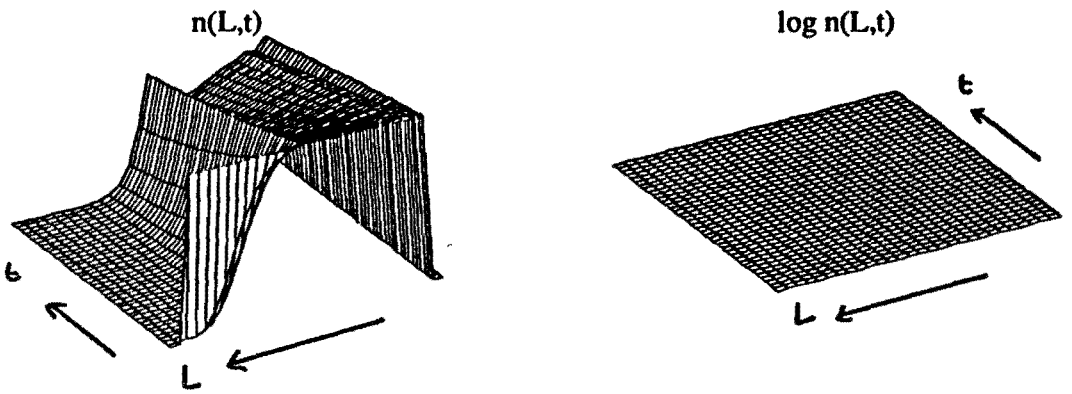



Figure 5: The test problem for the implicit discretisation scheme




| | |
|---|---|
| tmax = 1000 | dt = 25 |
| Lmax = 0.0001 | dL = 3.333e-006 |
| Lmin = 0 | max. peak = 0.4972 |
| L0 = 4.971e-005 | |
| n0 = 0.5 | |
| #L = 30 | |
| #t = 40 | |

Figure 6: Solution of Tavare's problem using an implicit discretization scheme

## 4.3 Reparametrisation

We assume for simplicity that every constant is 1. In harmony with general notation in PDE-courses $u$ and $x$ will be used in stead of $n$ and $L$. Consider again equation (7)

$$\frac{\partial u}{\partial t} + \frac{\partial G\, u}{\partial x} = 0 \tag{9}$$

where

$$G = G(t,x) = c(t) - e^{1/x} \tag{10}$$

and

$$c(t) = 2 - \int_0^\infty x^3 u(t,x)dx \tag{11}$$

The boundary conditions are

$$\begin{aligned} u_x(t,0) &= 0 \\ u(t,\infty) &= 0 \end{aligned}$$

and the initial value is

$$u(0,x) = u_0(x)$$

By introducing a new function

$$U(t,x) = \int_0^x x'^3 u(t,x')dx' \tag{12}$$

and noting that

$$u(t,x) = \frac{U_x(t,x)}{x^3} \tag{13}$$

one can rewrite (9) as

$$xGU_{xx} + xU_{xt} + (xG_x - 3G)U_x = 0 \tag{14}$$

and (11) as

$$c(t) = 2 - U(t,\infty) \tag{15}$$

By now the original model is reduced from a first order PDE with an extra integral equation to a second order PDE.

By assuming that $u(t,x)$ will become zero if $x$ is lager than $x_{max}$ we can reduce (15) even more to

$$c(t) = 2 - U(t,x_{max}) \tag{16}$$

but we will not use this. Introducing

$$z = \frac{1}{x}$$

and

$$G(t,x) = H(t,z)$$

one can write

$$U(t,x) = V(t,z)$$

Thus

$$\begin{aligned} U_x &= -z^2 V_z \\ U_{xt} &= -z^2 V_{zt} \\ U_{xx} &= z^4 V_{zz} + 2z^3 V_z \end{aligned}$$

and (14) will now become

$$Hz^2 V_{zz} - V_{zt} + (5zH + z^2 H_z)V_z = 0 \qquad (17)$$

where

$$\begin{aligned} H(t,z) &= 2 - V(t,0) - e^z \\ H_z(t,z) &= -e^z \end{aligned}$$

The initial value for $V$ is

$$V(0,z) = V_0(z) = \int_0^{1/z} z'^3 u_0(z')dz'$$

and the boundary values are

$$\begin{aligned} V_z(t,0) &= 0 \\ V_z(t,\infty) &= 0 \end{aligned}$$

Note that in (17) one is tempted to replace $V_z$ with a new function $W$, but this is not usefull since $V$ appears in the definition of $H$ and in the initial value.

By now one obtains a fairly regular equation which can be implemented rather easily. Note that computing the original $u(t,x)$ from $V(t,z)$ can easily be done by numerical differentiation and should not be a problem.

The proposed procedure to solve this could be a discretisation. Because the equation (17) contains a mixed derivative, the procedure will be implicit and therefore hard to solve.

The second alternative is to propose an iterative method.

One starts out with proposing a value for $V(t,0)$ an then solve equation (17) which can be simplified by saying $W = V_z$, with a discretisation method.

Now $V(0,t)$ can be updated to a new value and the iteration can start again, but we will not follow this procedure but instead use the method of characteristics (Which is far better, as we discovered).

## 4.4 Characteristic method

We have to investigate the PDE (7) in which we have set all the constants equal to 1:

$$u_t + (Gu)_x = 0$$

where

$$G(t,x) = c(t) - e^{\frac{1}{x}}$$

and

$$c(t) = 2 - \int_0^\infty x^3 u(t,x)dx \tag{18}$$

This is a mass-balance equation and these kind of equations have the nice property that

$$\int_0^\infty u_0(x)dx = \int_0^\infty u(t,x)dx \quad \forall t$$

This is an important property, since it may tell us whether a numerical approximation is still valid or not.

We start with a rectangular initial function:

$$u_0(x) = \begin{cases} 0 & \text{if } x < x_{min} \\ \hat{u} & \text{if } x_{min} \leq x \leq x_{max} \\ 0 & \text{if } x > x_{max} \end{cases}$$

where $\hat{u}$ can be chosen. By specifying the value of $\hat{u}$ we can shift the curve $G = 0$ to the left or to the right. For instance, if we would like this curve to be 'in the middle' of our rectangular initial data, we could take

$$\hat{u} = 4\frac{2 - e^{\frac{2}{x_{min}+x_{max}}}}{x_{max}^4 - x_{min}^4}$$

Recall that a discontinuity in the initial value will propagate along a characteristic curve and thus we will not have to bother with it, since we will rewrite the system into characteristic ODE's.

The characteristic equations are by now:

$$\begin{aligned} \dot{t}^s &= 1 \\ \dot{x}^s &= G \\ \dot{u}^s &= -G_x u \end{aligned}$$

and the initial values are

$$\begin{aligned} t^s(0) &= 0 \\ x^s(0) &= s \\ u^s(0) &= u_0(s) \end{aligned}$$

By taking a finite interval on which $u_0$ is not equal to zero, the calculation of (18) is restricted to a bounded interval since the points in the $(x,t)$-plane which will be reached by a characteristic starting outside the nonzero initial $u_0$, will have value $u = 0$. Thus, and this is a major improvement, we can restrict ourselves to calculating the values of $u$ between the most left characteristic and the most right.

We also started with an interval which does not contain 0, since by then we would be in trouble if we wanted to calculate $G(t,0)$.

How do we calculate $c(t)$? We can easily compute the integral of $u(x,t)x^3$ if the $u$ is given. And because the interval on which $u$ is non-zero is bounded, this can be done rather accurate. But, by this procedure, $c(t)$ will be no more than a so called 'static' variable. It will not contain any new information about, for instance, the accuracy of our solution. This is a pity, since we start with a nice integral equation which c has to satisfy. Thus, one rewrites this equation to the next form, taking into account the original PDE

$$\dot{c} = -3c \int_0^\infty z^2 u(t,z)dz + 3 \int_0^\infty z^2 u(t,z) e^{\frac{1}{z}} dz$$

The main idea for our method is to calculate all the characteristics at once. This is not an easy task, since there are infinitely many of them. Thus we will have to make a discretisation at the initial value to avoid this.

By now we are able to formulate our method:

$$
\begin{aligned}
\dot{x_0} &= G(t, x_0) \\
\ldots &= \ldots \\
\dot{x_n} &= G(t, x_n) \\
\dot{u_0} &= -G_x(t, x_0)u_0 \\
\ldots &= \ldots \\
\dot{u_n} &= -G_x(t, x_n)u_n \\
\dot{c} &= -3c\int_0^\infty z^2 u(t,z)dz + 3\int_0^\infty z^2 u(t,z)e^{\frac{1}{z}}dz
\end{aligned}
$$

This is a $2n+3$ dimensional ODE system in which we only have to specify our initial data. Well, these are

$$
\begin{aligned}
x_0(0) &= x_{min} \\
x_1(0) &= x_{min} + \delta_x \\
\ldots &= \ldots \\
x_n(0) &= x_{min} + n\delta_x = x_{max} \\
u_0(0) &= u_0(x_0) \\
\ldots &= \ldots \\
u_n(0) &= u_0(x_n) \\
c(0) &= 2 - \int_0^\infty x^3 u_0(x)dx
\end{aligned}
$$

where $\delta_x$ is the step-size.

This can now be calculated with Matlab, and thus we have done. From figure (7) we can see the following things:

1 We started with the initial non-zero $U_0$ at the right of the line $G = 0$.

2 The characteristic lines turn to the right at first.

3 The line $G = 0$ turns to the right too, but faster than the characteristic lines.

4 Thus there will be a point when some characteristics are 'overcome' with the line $G = 0$. They will have to go to the left by now.

5 The characteristic most on the left will eventually get a derivative in the minus x-direction of infinity. Thus it will not move in the t-direction anymore. By now, our program breaks down and reports a singularity.

6 From the curve for $c(t)$ we see that the $c(t)$ value will decrease in time. This is not clear up front, since it decreases only slightly and it takes a t-value of over one hundred for a 30% reduction

7 The $u(x)/\hat{u}$-curve tells us that the value of $u$ is decreasing in time. Denote also that the characteristic which provides the singularity carries very little weight and will carry even less weight in the calculation of $c(t)$, since it will be multiplied with a third power of $x$. An improvement of the method will thus be to skip the troublesome characteristic, but this goes beyond our investigations.

8 One of the main questions is by now : Why does the method break down? This is not an easy question. By looking again at the characteristic equations

$$\begin{aligned} \dot{t} &= 1 \\ \dot{x} &= c(t) - e^{\frac{1}{x}} \\ \dot{u} &= -\frac{1}{x^2} e^{\frac{1}{x}} u \end{aligned}$$

we see that

8.1 $u$ decreases if $t$ increases,

8.2 and so does $x$, if we are on the left of $G = 0$,

8.3 thus if we are on the left of $G = 0$, we expect that $u$ will tend to zero and the more we are on the left, the faster it will be.

By substituting the values for $x$ and $u$ we see that the order of decreament is way larger than the order of increament in $t$. Thus we would expect that there will be a horizontal part in the characteristic curves. This does not coincide with our theory, since we would expect that the process could run forever. So, what's wrong? Basic algebra shows that on the left of $G = 0$ the x-characteristics have negative second derivatives:

$$\begin{aligned} \ddot{x} &= \dot{c} + \frac{1}{x^2} e^{\frac{1}{x}} \dot{x} \\ &= \int_0^\infty z^3 G_z u dz + (\frac{1}{x^2} e^{\frac{1}{x}}) G \end{aligned}$$

since

$$\dot{c} = \int_0^\infty z^3 G_z u dz$$

and both terms are negative. Thus, they will not go 'up' again unless they are overcome with the line $G = 0$. The physical meaning will be that some crystals will dissolve in finite time. And this is not unlikely, since we do not stir our coffee for ever and ever.

9 The curve $G = 0$ plays an important role in the $(x, t)$-plane. It will try to follow the most to the right situated characteristic, thus exterminating all other characteristis. (It's almost human)

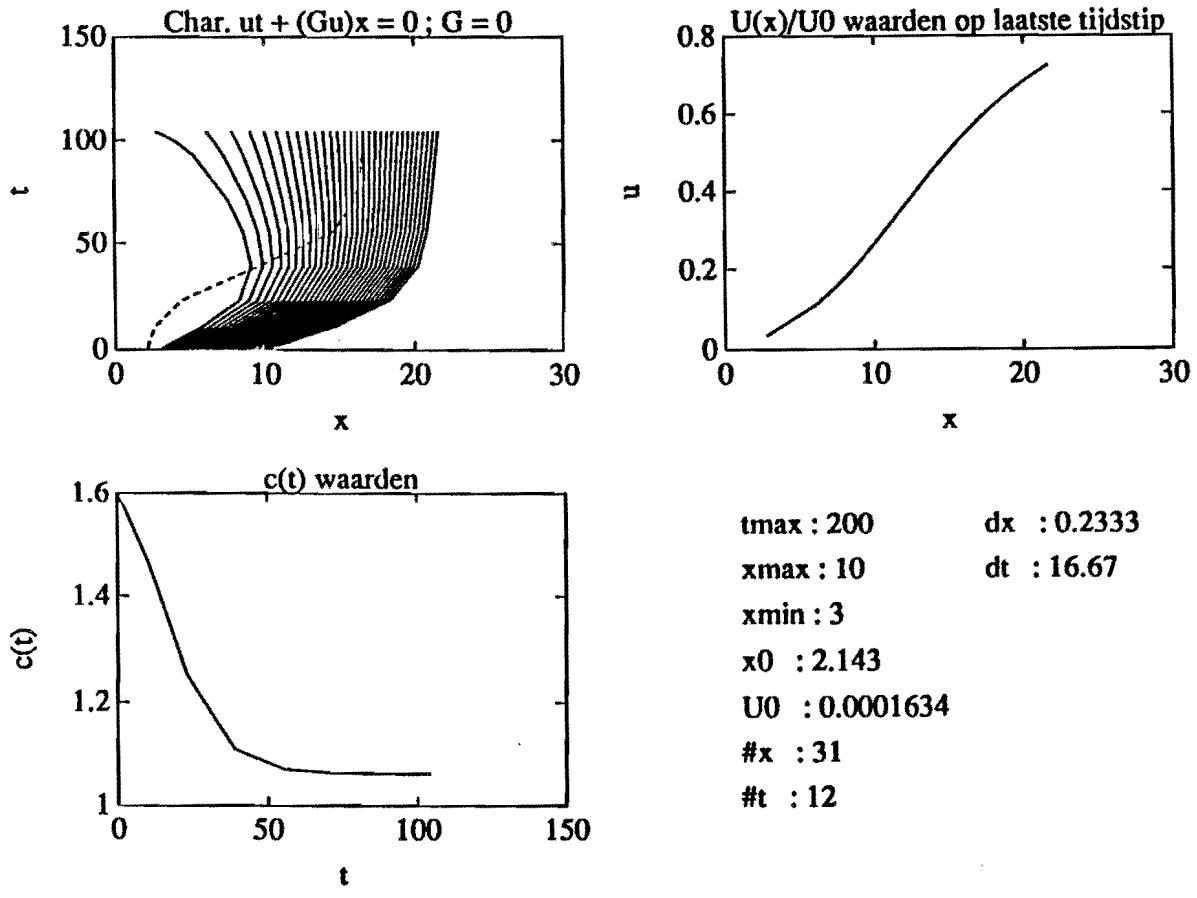



Figure 7: A typical example, Rectangular initial data

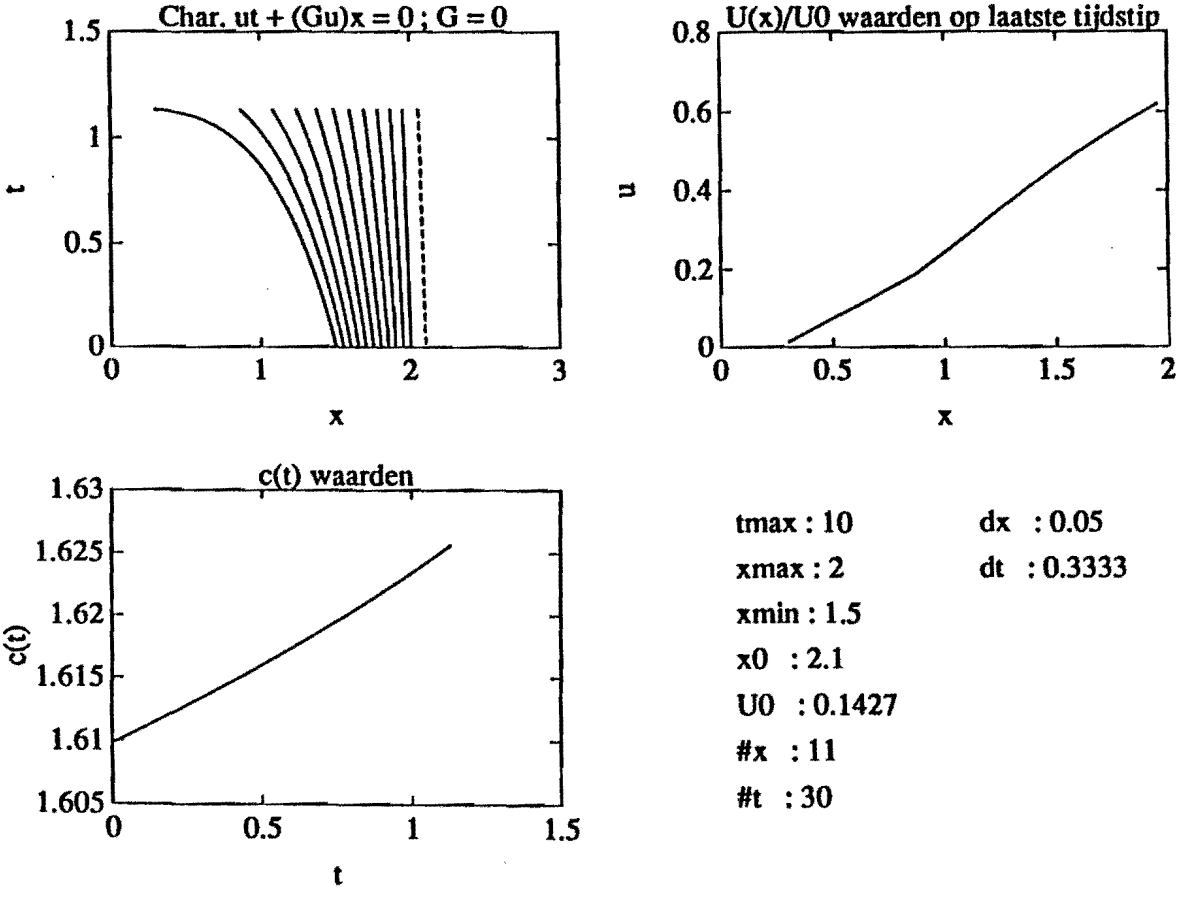



Figure 8: Another typical example, Rectangular initial data

For comparison reasons, we have also looked at the problem with initial data

$$U_0(x) = \begin{cases} 0 & \text{if } x < x_{min} \\ a(x - x_{min}) & \text{if } x_{min} \leq x \leq \frac{x_{min}+x_{max}}{2} \\ a(x_{max} - x) & \text{if } \frac{x_{min}+x_{max}}{2} \leq x \leq x_{max} \\ 0 & \text{if } x > x_{max} \end{cases}$$

See figure (9).

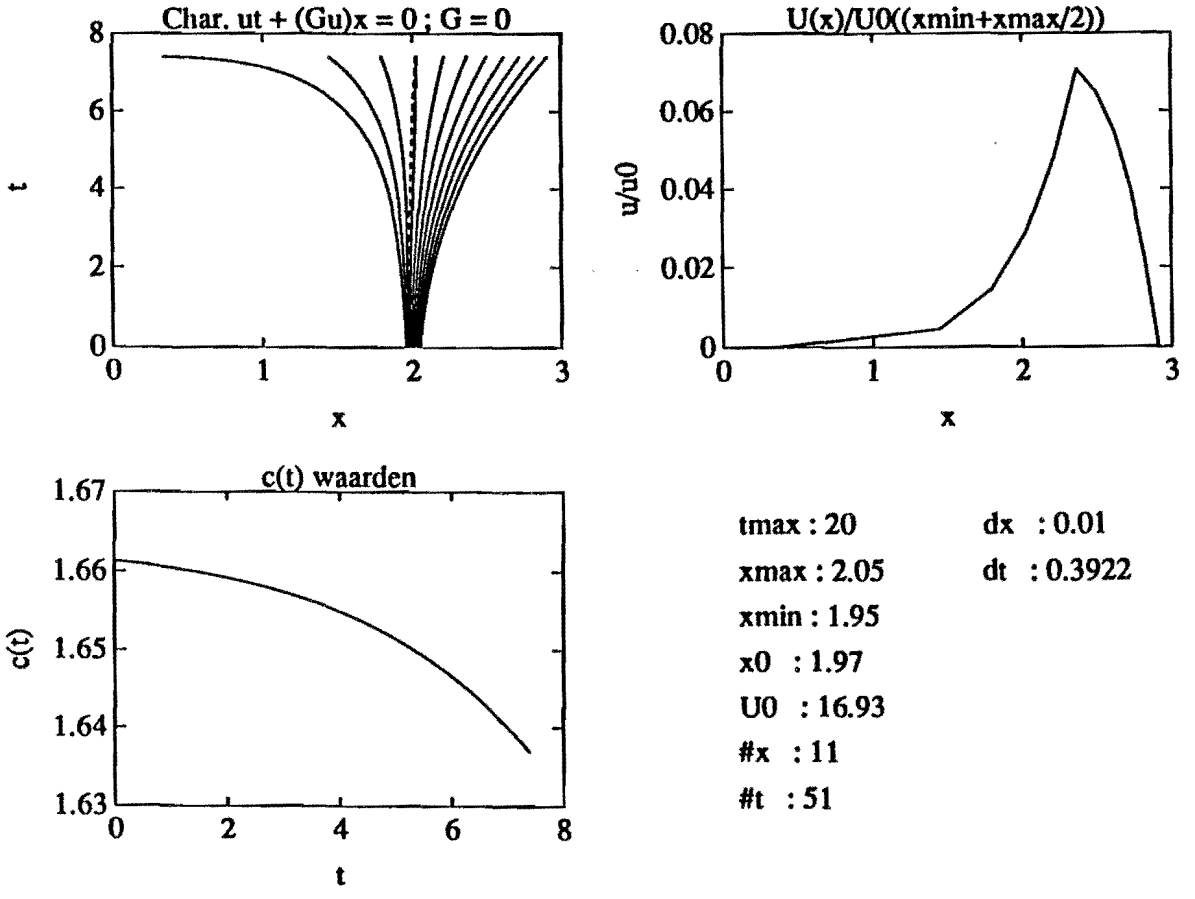



Figure 9: Another typical example, Triangular initial data

# 5 Conclusions

Referring to the first section of this article, we can formulate our starting point. It was to gain insight in the process of crystal growth. In spite of this short formulation the problem appeared to be quite complicated. So we only studied a special case of crystal precipitation: the Ostwald ripening.

The rather simple looking model of Tavare was approached by four methods. Reason for the need of several methods succeeding eachother lies in a circular dependency. The first-order PDE in two variables seems regular but contains a hidden integral depending on the differentiated variable. To get round the integral the PDE first was tackled with an explicit discretisation scheme. Although of order $h^2$ it didn't satisfy due to the numerical diffusion and dissipation. The diffusion also disturbed the method with implicit discretisation. The problem with diffusion was solved by a reparametrisation of the variables, for the first-order PDE became a second-order equation. But we didn't work with this extensive approach.

In fact the final method offered more results. Based on the characteristics of the PDE, the equation was reduced to a set of ODE's, including the integral. The restriction is in the number of characteristics to be followed. The basic physical aspects of Ostwald ripening are recognised, despite of the explanation of the speed of the growth for small crystals.

Nevertheless, a proceeding study will be valuable. First of all the computations are made with the limitations of an PC-AT, meaning a restricted number of steps. Furthermore all the constants in the problem are taken equal to 1. This doesn't alter the problem really, except for the power coefficients maybe. At last there may be an extension of the method by excluding horizontal characteristics out of the model. Apparently the at least the numerics are far away from using them to make highquality photographs, so a continuing research surely is possible.

# References

[Cas87] P.E. Castro et alii
Conservation Laws in Crystal Precipitation, 1987
pp. 32-39

[Tav87] N.S. Tavare
Simulation of Ostwald ripening in a Reactive Batch Crystallizer
American Institute of Chemical Engineers Journal, vol.33, 1987
pp. 152-156

# A MATHEMATICAL MODEL FOR THERMAL IMAGING

# Contents

# 1 Introduction

Thermal imaging is a rather complex process. An image is made of a thin film consisting of a donor and a receptor. Both are polymers or mixtures of polymers. The donor contains a certain amount of dye, which is a colorant. The donor is heated by a heat source, mostly a thermohead consisting of little semiconductor elements. One wants to give the receptor a certain colour. By putting heat into the donor, the dye comes free and flows into the receptor. In the receptor it's hard for the dye to move, as if it sticks to the receptor molecules.

The process of transport of dye is a sort of diffusion process, but one can prove that it cannot be a linear one. This is logical, because the dye is not completely free to flow around in the receptor.

When you increase the time that the donor is heated, more dye will flow into the receptor. The initial dye concentration in the donor may also be an important factor for the colour of the receptor. People who work with thermal imaging may be interested in questions like:

"How long should I heat the donor until the concentration of dye in the receptor is high enough ?", or :

"What initial dye concentration should we have, to get a good image ?"

For these and other questions the concentration of dye as a function of time and position is very important. This concentration will be the subject of the next sections.

# 2 Stating the problem

We want to know the concentration distribution of the dye on a certain moment in time. Important factors influencing this distribution may be :

- the material characteristics of the donor, receptor and dye
- the initial concentration of dye in the donor
- the process of heating until that moment
- the length, width and thickness of the donor and receptor.

We will describe the position of a dye particle in the donor or receptor by a positive real number, which expresses the distance from the heat source. This means, that we take a one-dimensional model. We can do this, because we assume, that the donor and receptor are homogeneous bodies. Furthermore the lengths of donor and receptor are very big compared with their thicknesses: the thickness of the donor was given to be $5\mu$ and the thickness of the receptor is equal to 45 $\mu$, while the lengths can be assumed of order inches.

We don't know for which concentration of dye the image will have a high quality. That will be left to the opinion of people working with thermal imaging.

From the experiments described in the article we get some data about the so-called "half-life" of the process. This is the time it takes half of the amount of dye to transfer to the receptor. In the same way the "quarter-life" is defined as the time it takes a quarter of the dye to flow to the receptor. Two half-lives and two quarter-lives were measured, with different initial dye concentrations and donor thicknesses.

| Concentration | Thickness | Half-life | Quarter-life |
|---|---|---|---|
| 0.25 | $1.5\mu$ | 39 $s$ | |
| 0.125 | $3.5\mu$ | 60 $s$ | |
| 0.0625 | $9.0\mu$ | | 38 $s$ |
| 0.03125 | $17.5\mu$ | | 150 $s$ |

Table 1: Experimental data for half- and quarter-lives.

We also know from experiments that, for every fixed time, the concentration distribution has a so-called "knee-shape", like the right hand side of a Gaussian curve.

Now we can describe our problem as follows. We want to find a concentration function, depending on time and position, that :

- will give approximately the same half- and quarter-lives as found with the experiments
- has the knee-shape for every fixed time $t$.

We will describe a model for the diffusion process. Then we will calculate a solution that fits the data, using numerical methods.

# 3 Previous models.

Two models have already been tested in the past: a model for linear diffusion and an alternative approach, with a power function. We use the following variables and parameters to describe those models:

**Variables:**

| name | symbol | (unit) |
|---|---|---|
| Concentration of the dye | : $u$ | : $(mol/m)$ |
| Time | : $t$ | : $(s)$ |
| depth | : $x$ | : $(m)$ |

**Parameters:**

| name | symbol | (unit) |
|---|---|---|
| thickness of the donor | : $a$ | : $(m)$ |
| thickness of the receptor | : $b$ | : $(m)$ |
| Initial concentration of the dye | : $C$ | : $(mol/m)$ |
| Diffusivity | : $D$ | : $(m^2/s)$ |

The following model for linear diffusion is inconsistent with the experimental data:

$$\frac{\partial u}{\partial t} = D\frac{\partial^2 u}{\partial x^2}, \quad 0 < x < b, t > 0 \tag{1}$$

$$\frac{\partial u}{\partial x}(0,t) = 0, \quad t > 0 \tag{2}$$

$$\frac{\partial u}{\partial x}(b,t) = 0, \quad t > 0 \tag{3}$$

with initial condition:

$$u(x,0) = \begin{cases} C, & 0 \leq x \leq a \\ 0, & a < x < b \end{cases} \tag{4}$$

One can prove that if the diffusion is linear, the half- and quarter-lives don't depend on the initial concentration of the dye and the thickness of the donor, and this is not what we want.

The next model can be solved analytically:

$$u_t = \frac{\partial^2(u^m)}{\partial x^2}, \quad -b < x < b, t > 0, m > 0 \tag{5}$$

$$u(x,0) = \begin{cases} C, & |x| < a \\ 0, & a < |x| < b \end{cases} \tag{6}$$

$$\frac{\partial u}{\partial x}(-b,t) = \frac{\partial u}{\partial x}(b,t) = 0 \tag{7}$$

It can be proved that this model has a unique solution but it is unclear if the solution is consistent with the experimental data.

We will search for a new model that does meet with the experiments. This new model will be described in the sequel.

# 4 Assumptions.

First we saw that the donor and the receptor needed to be treated differently. First of all, the materials that the donor and the receptor are made of, are different. This gave us the idea to try for each a different diffusion equation. Moreover, since the main interest of the problem lies in the concentration of the dye in the receptor, we decided that the donor might best be modelled by linear diffusion.
The experiments showed that the receptor definitely could not be treated that way. Therefore we focused on finding different diffusion coefficients for the receptor which could be defended physically and which would meet with the experiments.

We assumed the following:

- The donor and the receptor are both homogeneous media.
- We can use a one-dimensional model, as we said earlier.
- There is linear diffusion in the donor.
- In the receptor we have non-linear diffusion, with a diffusion coefficient depending on the concentration of the dye in the receptor.
- There is no influence of the temperature on the diffusion process. This means that the diffusion coefficients don't depend on the temperature.

In the following section we will introduce the equations that describe this process.

# 5 Mathematical model.

First we expand the domain of the position $x$ to the complete set of real numbers (between [-b,b]) by mirroring the initial concentration function in the point $x = 0$.

We divide the $x$-axis into two parts:

- part I: $x \in [-a, a]$
- part II: $x \in [-b, -a] \cup [a, b]$

We define the concentration in each part by a partial differential equation (PDE) and then make a coupling on the boundary of the two parts.

We use the following variables and parameters in the model:

**Variables:**

| name | symbol | (unit) |
|---|---|---|
| Concentration of dye in donor | : $v$ | : $(mol/m)$ |
| Concentration of dye in receptor | : $u$ | : $(mol/m)$ |
| Position | : $x$ | : $(m)$ |
| Time | : $t$ | : $(s)$ |

**Parameters:**

| name | symbol | (unit) |
|---|---|---|
| Thickness of donor | : $a$ | : $(m)$ |
| Thickness of donor plus receptor | : $b$ | : $(m)$ |
| Diffusion constant in donor | : $D_0$ | : $(m^2/s)$ |
| Initial dye concentration donor | : $u_0$ | : $(mol/m)$ |
| Diffusion function in receptor | : $D$ | : $(m^2/s)$ |
| Constant in diffusion function | : $\alpha$ | : $(m^3/(mol \cdot s))$ |
| Constant in diffusion function | : $\beta$ | : $(m^2/s)$ |

Moreover, we denote
$(.)_x := \frac{\partial(.)}{\partial x}$ and $(.)_{xx} := \frac{\partial^2(.)}{\partial x^2}$.

The concentration $v$ in part I will be expressed through a linear diffusion equation, as follows:

$$v_t = D_0 v_{xx}, \; -a \leq x \leq a \tag{8}$$

$$v(x,0) = v_0, \; -a \leq x \leq a \tag{9}$$

The concentration in part II is expressed through a non-linear diffusion equation, often used in literature:

$$\begin{array}{lcll} u_t & = & (D(u)u_x)_x, & -b < x < b, t > 0 \\ u_x(-b,t) & = & 0, & t > 0 \\ u_x(b,t) & = & 0, & t > 0 \\ u(x,0) & = & 0, & -b < x < b \end{array}$$

where we remark that we can rewrite the first line:

$$u_t = D(u)u_{xx} + D_u(u)(u_x)^2$$

The first and second boundary condition express isolation. $D(u)$ is a suitable function of the concentration, called the diffusion coefficient.
To keep the computations simple, we try for $D$ a linear function of the concentration:

$$D(u) = \alpha \cdot u + \beta \tag{10}$$

Then $\alpha$ has to be negative and $\beta$ has to be positive, because it is logical that if the concentration is higher, the diffusion process will be slower. If we can not find any $\alpha$ and $\beta$ such that the experimental data are fit, we will have to try an other function for $D$.

The physical coupling of the two parts can be found using the law of conservation of mass and it is expressed by the following conditions:

$$\frac{\partial v(-a,t)}{\partial x} = \frac{\partial u(-a,t)}{\partial x} \tag{11}$$

$$\frac{\partial v(a,t)}{\partial x} = \frac{\partial u(a,t)}{\partial x} \tag{12}$$

On time zero, both sides of the equations stated above are equal to zero, which means that nothing will happen to start the process of diffusion. Therefore we

need some extra conditions to make the process start. Various equations can be used to fulfill this problem. We chose to take:

$$v(-a,0) = u(-a,0) \tag{13}$$

$$v(a,0) = u(a,0). \tag{14}$$

After we have found a solution to the concentration we can look at the half- and quarter-lives of our solution. This means that we search for the point in time where half (or a quarter) of the amount of dye has flown into the receptor. The exact amount at time $t$ is:

$$\int_a^b u(x,t)dx \ .$$

# 6 Numerical Methods.

Because the partial differential equations introduced in the previous section can not be easily solved analytically, we decided to solve them numerically, by an explicit method.

To reduce computation time, we first skip the mirroring of the $x$-axis in $x = 0$ again and introduce the extra condition:

$$\frac{\partial v}{\partial x}(0,t) = 0 \tag{15}$$

Furthermore, we use the following notations:
Define a grid on the time axis of width $\Delta t$ and number of points $M$ and a grid on the $x$-axis of width $\Delta x$ and number of points $N$.
$n_a = 0.1 \cdot N$ is defined as the gridpoint on the $x$-axis that represents the position $x = a$. (We used the fact that $a = 0.1 \cdot b$.)
We define:

$$x_i := i \cdot \Delta x,\ 0 \leq i \leq N \tag{16}$$

$$t_j := j \cdot \Delta t,\ 0 \leq j \leq M \tag{17}$$

and,

$$u_i^j = \begin{cases} v(x_i, t_j), & 0 \leq i \leq n_a,\ 0 \leq j \leq M \\ u(x_i, t_j), & n_a + 1 \leq i \leq N,\ 0 \leq j \leq M \end{cases} \tag{18}$$

From literature we know that for stability of an explicit scheme for the linear diffusion equation in the donor, $v_t = D_0 v_{xx}$, the following condition is necessary:

$$\frac{D_0 \cdot \Delta t}{(\Delta x)^2} < \frac{1}{2}, \tag{19}$$

For stability of a scheme for the non-linear PDE that describes the diffusion in the receptor, $u_t = (D(u)u_x)_x$, we will use the same condition, but with $D_0$ replaced by the maximal value of the function $D(u) = \alpha \cdot u + \beta$. Since the concentration $u$ is at least zero and $\alpha$ is smaller than zero, this maximal value is equal to $\beta$.

When we have chosen a value for the number of gridpoints in the $x$-direction, $N$, the timestep $\Delta t$ can be determined by the following equation:

$$\frac{max(D_0, \beta) \cdot \Delta t}{(\Delta x)^2} < \frac{1}{2}, \tag{20}$$

Now we can introduce the numerical schemes.

For the donor part ($1 \leq i \leq n_a,\ j \geq 0$), we use:

$$\frac{u_i^{j+1} - u_i^j}{\Delta t} = D_0 \cdot \frac{u_{i+1}^j - 2u_i^j + u_{i-1}^j}{(\Delta x)^2} \tag{21}$$

with on $i = 0$:

$$\frac{u_0^{j+1} - u_0^j}{\Delta t} = 2D_0 \frac{u_1^j - u_0^j}{(\Delta x)^2} \tag{22}$$

For the receptor part ($n_a + 1 \leq i \leq N - 1,\ j \geq 0$), we use:

$$\frac{u_i^{j+1} - u_i^j}{\Delta t} = \beta \cdot \frac{u_{i-1}^j - 2u_i^j + u_{i+1}^j}{(\Delta x)^2} + \frac{1}{2}\alpha \cdot \frac{(u_{i-1}^j)^2 - 2(u_i^j)^2 + (u_{i+1}^j)^2}{(\Delta x)^2} \tag{23}$$

The scheme that takes care of the coupling of the two previous schemes ($j \geq 0$) is:

$$\frac{u_{n_a}^{j+1} - u_{n_a}^j}{\Delta t} = D_0 \cdot \frac{2u_{n_a-1}^j - 5u_{n_a}^j + 4u_{n_a+1}^j - u_{n_a+2}^j}{(\Delta x)^2} \tag{24}$$

while boundary conditions ($j \geq 0$) are described by:

$$\frac{u_N^{j+1} - u_N^j}{\Delta t} = 2\beta \cdot \frac{u_{N-1}^j - u_N^j}{(\Delta x)^2} + \alpha \cdot \frac{(u_{N-1}^j)^2 - (u_N^j)^2}{(\Delta x)^2} \tag{25}$$

# 7 Results

We wrote a computer program that can compute for any value of the constant parameters (if permitted: for example $\alpha$ should be $< 0$, $\beta$ should be $> 0$) the concentration of the dye on any point of the $x$-axis at any time point.

We find that for $N$ big, for example $N = 1000$, the computation time is too big for the computer we use, to get an answer as fast as we like it to have. Therefore, we usually take $N = 200$, although for more accurate answers this value should be increased.

We runned the program for different values to the rest of the parameters, for example:

$$\begin{aligned} D_0 &= 1.0 \\ \alpha &= -0.5 \\ \beta &= 1.0 \end{aligned}$$

After computation of the numerical approximation for the concentration we can look at the half- and quarter-lives of our solution. We already introduced the equation to compute these values. The exact amount of dye flown into the receptor at time $t_j$ is:

$$\int_a^b u(x, t_j)dx \ .$$

We approximate this integral by a Riemann-sum:

$$\sum_{i=n_a+1}^{N} u_i^j \Delta x.$$

When we compute the half-lives and quarter-lives for every initial concentration and thickness of the donor as given in the table, we can compute a total relative error with the following formula:

$$\sum_{i=1}^{4} \left| \frac{t_i - t_i^*}{t_i^*} \right| \tag{26}$$

where $t_1^*$ denotes the first half-life mentioned in the table, $t_2^*$ denotes the second one, $t_3^*$ denotes the first quarter-life mentioned in the table and $t_4^*$ denotes the

second one. $t_1$, $t_2$, $t_3$ and $t_4$ denote their respective values as we compute them with our program.

For the example mentioned above, this total relative error appears to be 332%. Of course this is too big. We want to find better values to the parameters to reduce this error.

First, with a fixed value to $\alpha$, we runned the program to find a suitable value to $D_0$ and $\beta$, which we took equal to each other. This value appeared to be 0.37. Then we fixed this value to $D_0$ and tried to find $\alpha$ and $\beta$ such that the total relative error reached a minimum.

We plot the value to this formula for several $\alpha$ and $\beta$ in the next figure:

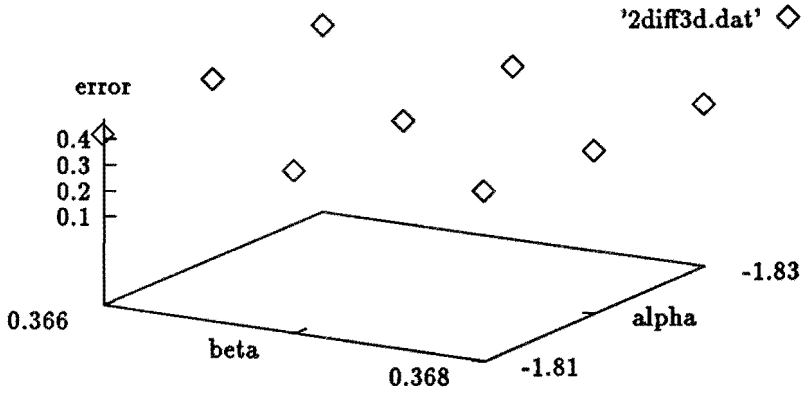



Figure 1: (26) as a function of $\alpha$ and $\beta$

The next values appear to give the best answer:

$$\begin{aligned} D_0 &= 0.37 \\ \alpha &= -1.83 \\ \beta &= 0.368 \end{aligned}$$

With a total relative error of 37,4%.

We present the values for the half- and quarter-lives for this $D_0$, $\alpha$ and $\beta$ in the following table: The values between brackets are the values obtained by the experiments, given by table 1.

| $D_0 = 0.37$, $\alpha = -1.83$ and $\beta = 0.368$ | | | |
|---|---|---|---|
| Concentration | Thickness | Half-life ($s$) | Quarter-life ($s$) |
| 0.25 | $1.5\mu$ | 39.4 (39) | 2.77 |
| 0.125 | $3.5\mu$ | 59.8 (60) | 9.21 |
| 0.0625 | $9.0\mu$ | 265 | 48.4 (38) |
| 0.03125 | $17.5\mu$ | 847 | 163 (150) |

Table 2. The best values found for $D_0$, $\alpha$ and $\beta$ and their half- and quarter-lives.

We see that the error of 37.4% is the result of an error of 1.4% for the half-lives and 36% for the quarter-lives.

We can plot the concentration as a function of $x$ on every fixed time $t$, for example on the just computed quarter- and half-lives:

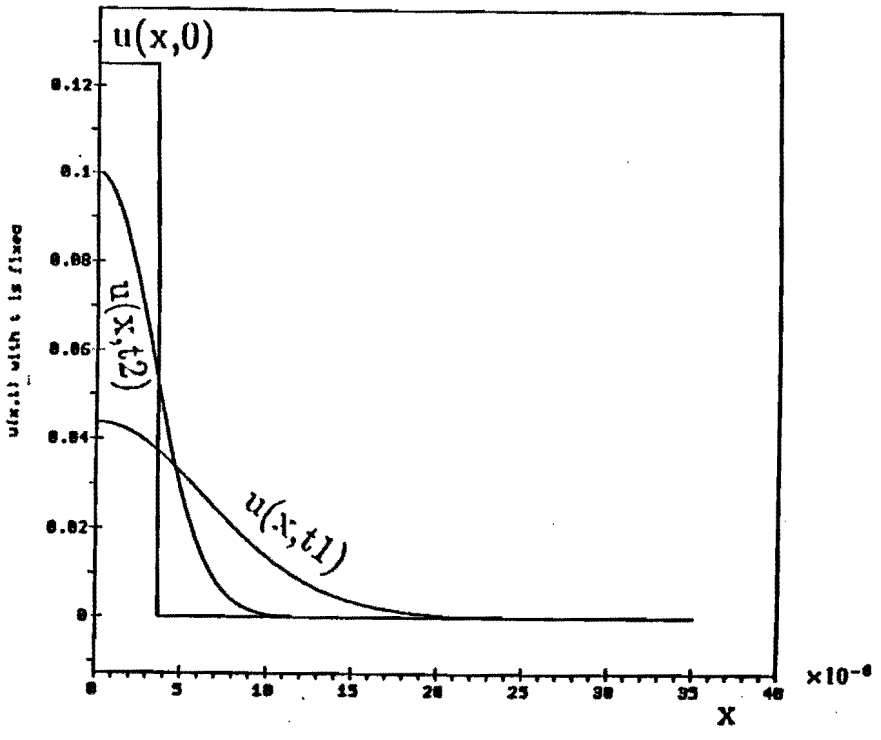



Figure 2: Diffusion half-lives (t1) and quarter-lives (t2)

We can see that the concentration has indeed the knee-shape that we want.

# 8 Conclusions and recommendations

In the previous section we saw that our choice for a linear diffusion coefficient gives a solution to the concentration of the dye in the receptor. The computed concentration as a function of its position on the $x$-axis seems to meet with the experiments when we see that on a fixed time point it has the wanted knee-shape.

When we look at the total relative error of the computed data, as introduced in the previous section, we see that the half-lives are reasonably fitted (there is an error of about 1.4%) but the quarter-lives are fitted with an error of 36%.

We can conclude that the computed values with the chosen set of constants do not meet exactly with the experimental values.

What can one do to improve this result?

First of all, one can run the program on a faster computer, such that more accurate grids can be chosen ($N$ can be increased). Moreover, one can play with the constant parameters and try to fit the data. The computer program in it present form produces answers very slowly, and it is quite hard a task to find fitting parameters values in a reasonably short time. It took quite a long time to fit the half-lives. To fit both half-lives and quarter lives will take much more time.

Thirdly, when one does not want to believe in a linear diffusion coefficient, one can try for instance a quadratic one or even more complicated ones. This yields that our numerical schemes can not be used anymore. Moreover, one even has to introduce a transformation of the parameters first to find a solution.

We believe that this all is beyond the scope of this project and leave it to further research.

# ADVECTION OF POLLUTED AIR

# Advection of Polluted Air

## Contents

# 1 A polluting factory

A factory of the international company Pilips emits as a consequence of its production process several more or less toxic substances. Due to wind and diffusion these pollutants are divided over a certain area. The city of Eijndhoven lies inside this area.

There are rules in Eijndhoven that state that the concentration of several toxic substances may not exceed some given norm. The factory will only get a permission to produce if the concentration of toxic substances, due to their emission is below this given norm.

After the substances are emitted they can react with each other or with substances that are already in the atmosphere. In this way new pollutants can be formed. The substances can also desintegrate as a result of sunlight, the lower pressure or the different temperature in the atmosphere. All pollutants that are formed by these reactions must fullfill the restrictions on concentrations of toxic substances.

Pilips wants to have a prognostic air quality model to predict, given the restrictions on concentrations, the allowed emission under all possible circumstances.

# 2 Assumptions

As may be noted from above, the omitted pollution should in no case exceed certain norms. Consequently, we are looking at worst case scenario's, i.e. those conditions for which the concentrations of the chemicals in the measurement point are as high as possible. However, those conditions must remain realistic. As a result of assuming the worst case scenario, we can make some simplifications. In the assumptions listed below, we mark the worst case scenario simplifications with an asterix (*). The assumptions are:

1. The wind can be written as a (small) stochastic wind added to the mean wind. As a result the concentration of the species can be written as a (small) stochastic concentration added to the mean concentration. We are only interested in the mean part of the concentration;

2. The mean wind is always blowing from the source to the measurement point following a straight line in the horizontal plane. If the velocity has another direction the result would be that the concentration in the measurement point would be lower. The line is straight because we neglect

the effect of the rotation of the Earth on the stream lines (the Coriolis effect). This seems reasonable because the characteristic "working" distance of the Coriolis force is much larger than the distance we actually consider. Moreover, including the effect of the Coriolis forces would increase the length of the path from the source point to the measurement point. Consequently, the concentration at the measurement point would be lower; *

3. The mean wind is assumed to be constant in time; the stochastic wind is assumed to have an expectation of zero;

4. The reaction rate velocities are only functions of the concentrations, i.e. not from pressure, temperature, windvelocity etc;

5. The shape of the country is flat. It is remarked that mountains would have a large influence on the concentrations (positive as well as negative);

6. Time effects of the concentrations are neglectable. This seems to be a reasonable assumption if we assume that the factory works continuously over a "long" period of time. With "long" we mean with respect to the time scales of convection and diffusion. A long period would be for instance a working day. Then the middle of the day would be a suitable time to measure the concentration, because we are only interested in the maximum concentration and not in the concentration at the begin or the end of the working day. The output concentration of the factory is constant in time. We assume that the factory works continuously during working time. The time between switching on in the morning and switching off in the evening is much larger than the timescale for convection and diffusion; *

7. The air is incompressible. This assumption is reasonable if we only look at air in low layers;

8. Turbulent diffusion is much stronger than molecular diffusion. This is based on the mixing length theory (see [Seinfeld 86]). Turbulent diffusion in the direction of the wind is negligable compared to convection;

9. The source of polution is a point;

10. Diffusion is equal in every direction.

# 3 The mathematical formulation

## 3.1 The equations

The basic equation governing pollution of $p$ chemical species in 3 dimensions is a conservation of mass equation. This equation is a nonlinear parabolic partial differential equation and is according to [Seinfeld 86] given by

$$\frac{\partial c_i}{\partial t} + \nabla \cdot (\mathbf{u}c_i) = D_i \nabla^2 c_i + R_i(c_1, \ldots, c_p, T) + S_i(\mathbf{x}, t), \qquad (1)$$
$$t > 0, \quad i \in \{1, \ldots, p\},$$

where

- $c_i$ is the concentration of chemical specie $i$ (in $mol/m^3$);
- $\mathbf{u}$ is the prescribed air velocity field in three directions (in $m/s$);
- $D_i$ is the molecular diffusion coefficient of specie $i$ (in $m^2/s$);
- $R_i$ is the reaction term or the rate of chemical formation (or depletion) of specie $i$ (in $mol/m^3 s$);
- $S_i$ is the source term or the rate of addition of specie $i$ (in $mol/m^3 s$).

Further $t$ denotes time (in $s$), $T$ denotes temperature (in $K$) and $\mathbf{x}$ are the coordinates of a cartesian system (in $m$). When we take the assumptions into account we get the following specifications:

- Assumption 1 leads to $\mathbf{u} = \bar{\mathbf{u}} + \mathbf{u}'$ and $c_i =< c_i > + c_i'$ where $\bar{\mathbf{u}} = \frac{1}{\tau}\int_t^{t+\tau} \mathbf{u} ds$, $\tau$ is an arbitrary time interval, and where $< \mathbf{u}' >=< c_i' >= 0$;
- Assumptions 2 and 3 lead to $\bar{\mathbf{u}}(\mathbf{t}) = \bar{u}\mathbf{e_x}$, where $\bar{\mathbf{u}}$ is a scalar constant and $\mathbf{e_x}$ is the unit vector in the direction of the wind;
- Assumption 7 leads to $\nabla \cdot \bar{\mathbf{u}} = 0$;
- Assumption 8 leads to $< u_j' c_i' >= -k_j \frac{\partial < c_i >}{\partial x_j}$, with $j = 1, 2, 3$ and $i = 1, \ldots, p$;

- Assumption 6 leads to $\frac{\partial < c_i >}{\partial t} = 0$ for all $i$;
- Assumption 9 leads to $S_i = Q_i \delta(\mathbf{x} - \mathbf{x_0})$, where $Q_i$ is the polution velocity of specie $i$ (in $mol/s$).

The notation $< f >$ is used to denote that an average of $f$ over several measurements is taken.

## 3.2 The boundary conditions

For the boundary conditions we consider three possibilities:

**total absorption :**
If the pollution reaches the ground, the ground absorbs it entirely. The corresponding boundary condition is $< c_i(\mathbf{x}, t) >= 0$ at $z = 0, t > 0$. The problem can be modelled by adding an extra source $-Q_i \delta(\mathbf{x} - \mathbf{x_0}^*)$ to $S_i$ in the right hand side of equation (1), with $\mathbf{x}^* = (x_1, x_2, -x_3)$;

**total reflection :**
The pollution is not absorbed by the ground at all. In that case we have complete reflection at the ground. The corresponding boundary condition is $\frac{\partial < c_i(\mathbf{x}, t) >}{\partial z} = 0$ at $z = 0$, $t > 0$. The problem can be modelled by adding an extra source $Q_i \delta(\mathbf{x} - \mathbf{x_0}^*)$ to $S_i$ in the right hand side of equation (1);

**total reflection and inversion :**
The pollution is not only reflected at the ground but also at height $z = H$ ($H > h$, the emission height of the chimney). The corresponding boundary conditions are $\frac{\partial < c_i(\mathbf{x}, t) >}{\partial z} = 0$ at both $z = 0$ and $z = H$, $t > 0$. The problem can be modelled by adding extra sources $Q_i \delta(\mathbf{x} - \mathbf{x_0}^* + 2kH)$ and $Q_i \delta(\mathbf{x} - \mathbf{x_0} + 2kH)$, $k \in \mathbb{Z}$, to $S_i$ in the right hand side of equation (1).

In all cases, the solution has to satisfy the condition $< c_i(\mathbf{x}, t) >= 0$ when $|\mathbf{x}| \to \infty$, $z > 0$.

## 3.3 The reduced model

We concentrate on a pollution involving only one specie that will not react with the environment, thus $p = 1$ and $R_1 = 0$. Furter, we consider the source to be located at the point $\mathbf{x_0} = (0, 0, h)$. We substitute $\mathbf{u} = \bar{\mathbf{u}} + \mathbf{u}' = \bar{u}\mathbf{e_x} + \mathbf{u}'$ and $c_i = c =< c > +c'$ in formula (1). Taking the average value of this equation and recalling

$$< \mathbf{u}' >=< c' >= 0,$$

$$< \mathbf{u}_j' c' >= -k_j \frac{\partial < c >}{\partial x_j},$$

$$\nabla \cdot \bar{\mathbf{u}} = 0,$$

$$\frac{\partial < c >}{\partial t} = 0,$$

$$S_1 = Q\delta(\mathbf{x} - \mathbf{x_0}),$$

yields

$$u\frac{\partial < c >}{\partial x} = \sum_{j=1}^{3}(D + k_j)\frac{\partial^2 < c >}{\partial x_j^2} + Q\delta(\mathbf{x} - \mathbf{x_0}).$$

Because molecular diffusion is small compared to turbulent diffusion and both are small compared to convection, the above equation can be reduced to become

$$u\frac{\partial < c >}{\partial x} = k_y\frac{\partial^2 < c >}{\partial y^2} + k_z\frac{\partial^2 < c >}{\partial z^2} + Q\delta(\mathbf{x} - \mathbf{x_0}). \qquad (2)$$

This equation will be starting point for the mathematical work.

# 4 The solution

## 4.1 The formulas

In [Seinfeld 86] solutions for equation (2) are given for the three different boundary conditions. They are for

**total absorption:**

$$< c >= \frac{Q}{2\pi\bar{u}\sigma_y\sigma_z} e^{\left(-y^2/2\sigma_y^2\right)} \left[ e^{\left(-(z-h)^2/2\sigma_z^2\right)} + e^{\left(-(z+h)^2/2\sigma_z^2\right)} \right];$$

**total reflection:**

$$< c >= \frac{Q}{2\pi\bar{u}\sigma_y\sigma_z} e^{\left(-y^2/2\sigma_y^2\right)} \left[ e^{\left(-(z-h)^2/2\sigma_z^2\right)} - e^{\left(-(z+h)^2/2\sigma_z^2\right)} \right];$$

**total reflection and inversion:**

$$< c >= \frac{Q}{2\pi\bar{u}\sigma_y\sigma_z} e^{\left(-y^2/2\sigma_y^2\right)} \sum_{n=-\infty}^{\infty} \left[ e^{\left(-(z-h+2nH)^2/2\sigma_z^2\right)} - e^{\left(-(z+h+2nH)^2/2\sigma_z^2\right)} \right].$$

The deviations $\sigma_y$ and $\sigma_z$ are according to [Zeedijk 91] defined by

$$\frac{d\sigma_y^2}{dx} = 2\frac{k_y}{\bar{u}}, \quad \frac{d\sigma_z^2}{dx} = 2\frac{k_z}{\bar{u}}.$$

In practice, they can emperically be given by

$$\sigma_y = k \cdot x^l, \quad \sigma_z = a \cdot x^b,$$

with $a$, $b$, $k$ and $l \in \mathbb{R}$. Values for $a$, $b$, $k$ and $l$ are given in the next subsection.

We remark that the emperical formulas must stand in non-dimensional form. Although a derivation of the equations was not given, we assume that the variables were non-dimensionalized on variables with value 1.

## 4.2 Emperical values for the deviation

The values for $a$, $b$, $k$ and $l$ depend on the kind of weather. For chimneys higher than 100 metre [Zeedijk 91] defines four classes:

**unstable weather** , almost no wind nor clouds;

**normal weather** , normal wind and cloudiness;

**stable weather** , strong wind and many clouds;

**very stable weather** , the situation at night.

Note that very quiet weather is unstable. This means that the influence of turbulence in the $y$- and $z$-direction is big. With stable weather the plume is kept together for a long time.

The corresponding values for the parameters are:

| class | $a$ | $b$ | $k$ | $l$ |
|---|---|---|---|---|
| unstable | 0.411 | 0.907 | 0.40 | 0.91 |
| normal | 0.326 | 0.859 | 0.36 | 0.86 |
| stable | 0.233 | 0.776 | 0.32 | 0.78 |
| very stable | 0.062 | 0.709 | 0.31 | 0.71 |

In case we have total reflection we can give the maximum concentration and the place where this maximum occurs explicitly. In [Zeedijk 91] these maxima are given by

$$c_{max} = \frac{Q}{\bar{u}\pi ak}\left(\frac{e^1 bh^2}{a^2(b+l)}\right)^{-\frac{b+l}{2b}}, \quad x_{max} = \left(\frac{bh^2}{a^2(b+l)}\right)^{1/2b},$$

where $e$ is the number 2.717....

# 5 Results

## 5.1 Figures

With the solutions given in section 4 we can make clear what shape the concentration has as a function of $x$ under different circumstances by substituting realistic numbers. Then we find the following figures:

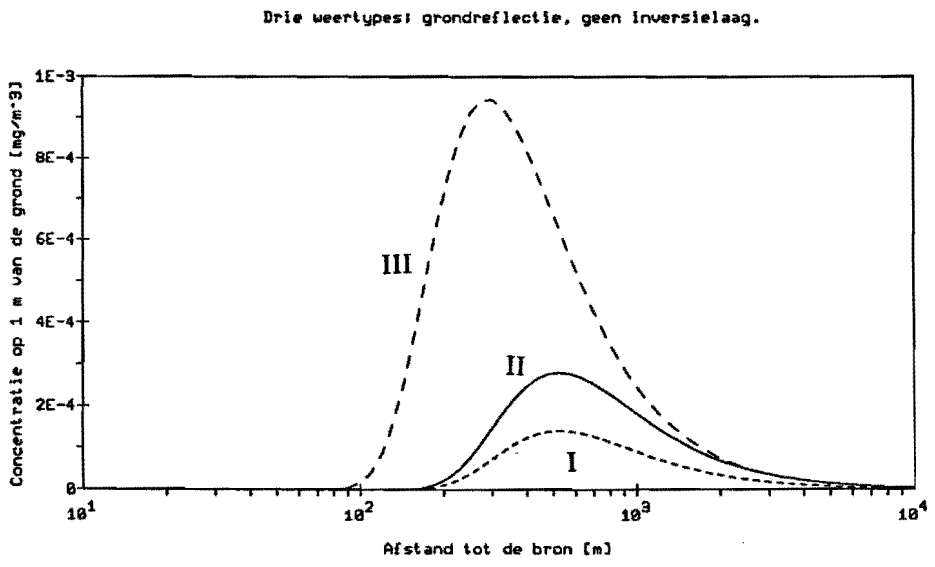



Figure 1: Reflection: influence of weather
I stable weather
II normal weather
III unstable weather .

Note the logarithmic scale

When there is almost no wind nor clouds the diffusion in the $z$-direction goes relatively fast. Therefore 1 metre above the ground the peak of the concentration will be higher and closer to the source, compared with normal weather.

At a distance bigger than 3 km of the source diffusion makes the concentration to become small.

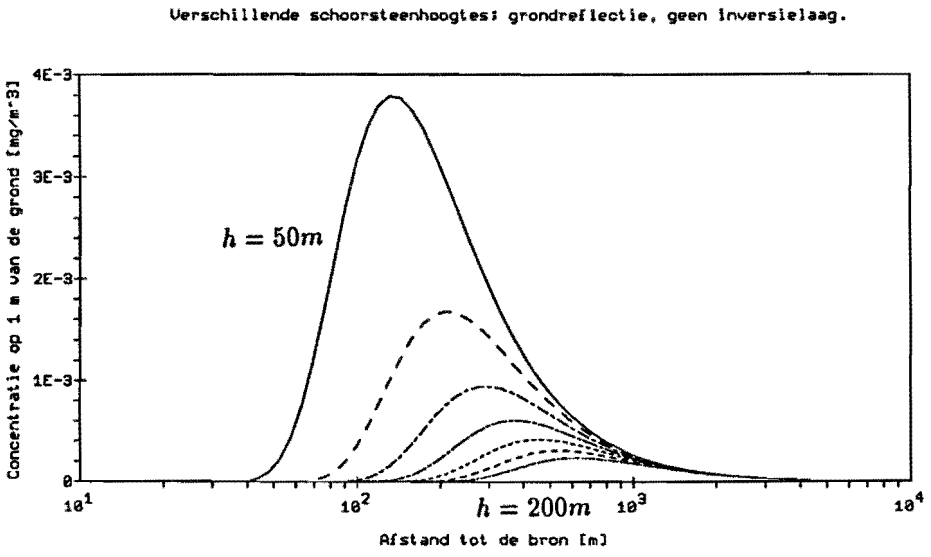



Figure 2: Reflection: influence of chimney height for h = 50 (25) 200 m.

The lower the chimney the higher and nearer the peak of the concentration will be. Note that the distance at which the concentrations starts to become bigger than 0 is more or less the same as the height of the chimney. At a distance bigger than 1 km the differences become negligible.

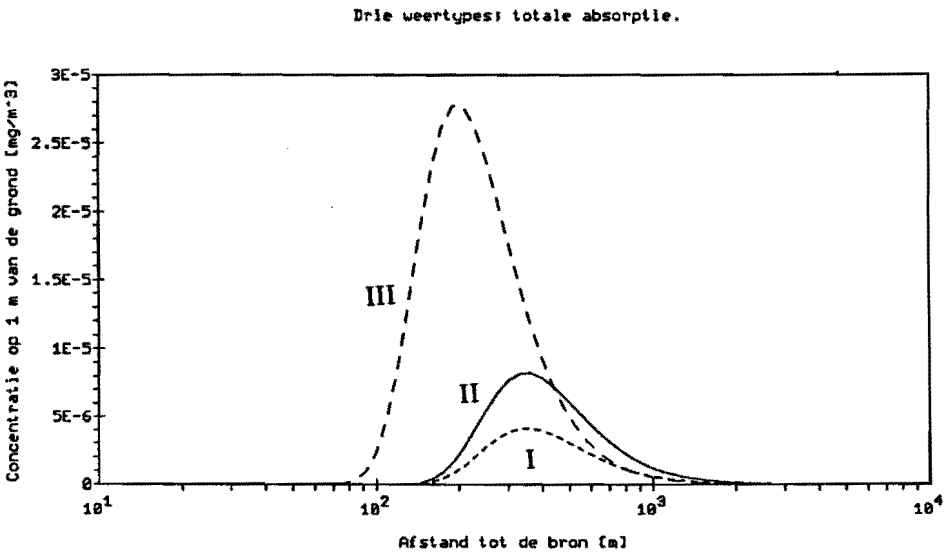



Figure 3: Absorption: influence of weather
I stable weather
II normal weather
III unstable weather.

Note the different scale on the $y$-axis.

Compared with reflection the peak is closer to the source, lower and smaller.

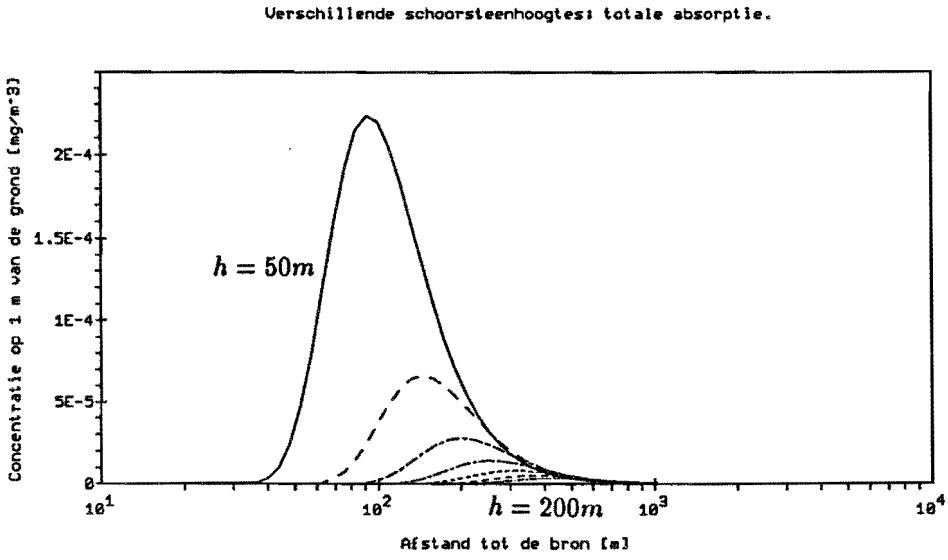



Figure 4: Absorption: influence of chimney height for h = 50 (25) 200 m.

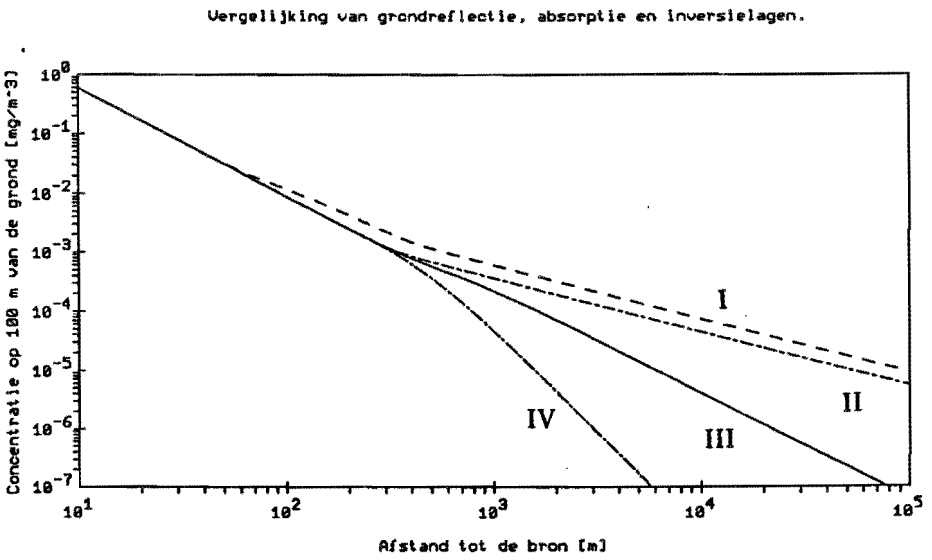



Figure 5: Concentration at height 100 m
I inversion at 120 m
II inversion at 200 m
III total reflection
IV total absorption

There is no difference between the different situations until a boundary layer is reached. With exception of the case that there is an inversion layer at height 120 m this layer is positioned at 100 m from the centre of the plume. That means that the place at which differences between the curves become visual must be the same for the three cases. This is shown in the figure.

Absorption bends the curve down, reflection moves it up and inversion bend it up. The curves of inversion layers on different height are parallel after bending.

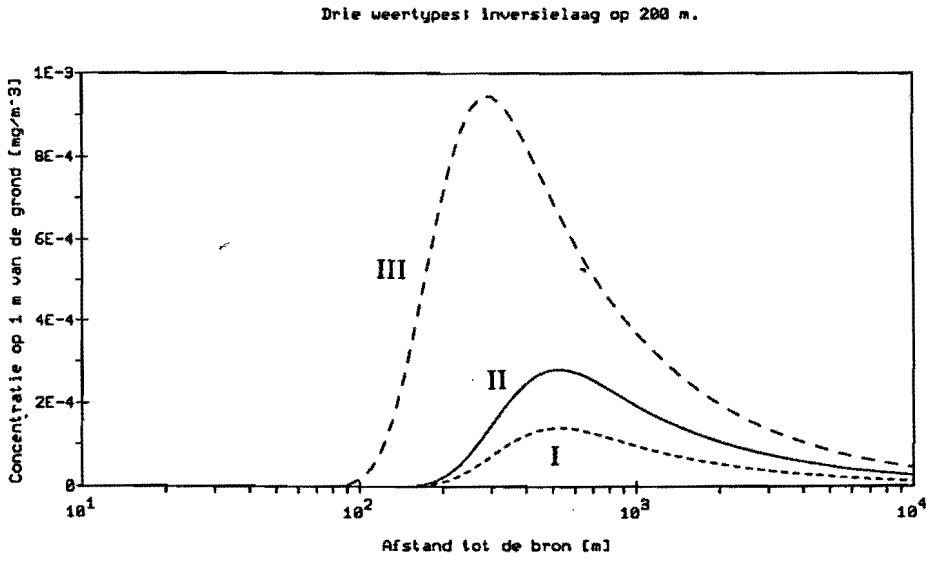



Figure 6: Inversion at 200 m: influence of weather
I stable weather
II normal weather
III unstable weather.

Compared with reflection the only difference is that the tails of the curves are thicker: the reflection is only noticable after the peak has occured.

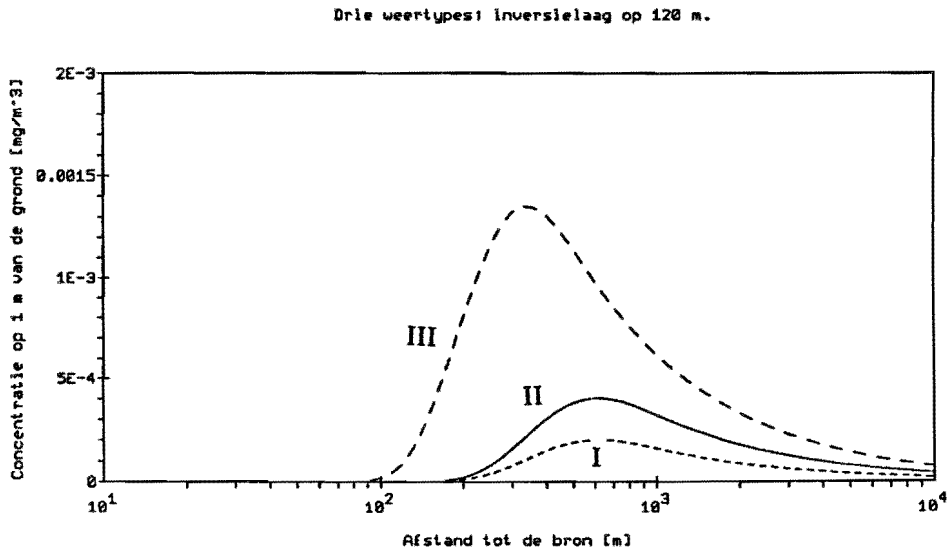



Figure 7: Inversion at 120 m: influence of weather
I stable weather
II normal weather
III unstable weather.

Now the reflection is also noticable at the peak itself. Therefore the curves are higher everywhere.

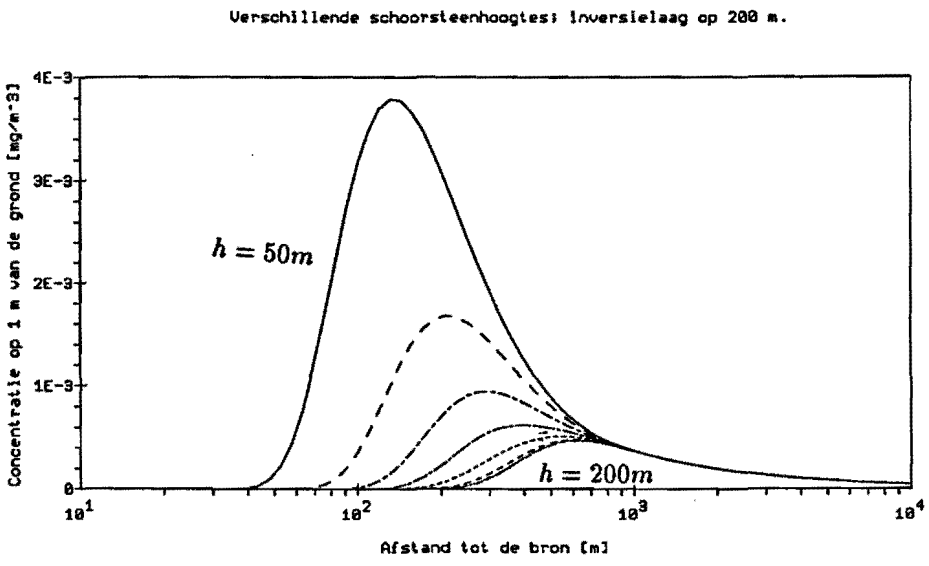



Figure 8: Inversion at 200 m: influence of chimney height for h = 50 (25) 200 m.

## 5.2 Calculations

To calculate the allowed emission under all possible circumstances for a given measure point at distance $x$ of the source, the following items must be done:

- Choose the worst case situation under realistic circumstances
- Use the corresponding figure in section 5.1 to estimate the maximum concentration $c(x)$ at the measure point if the emission velocity would be $Q = 100 \quad mg/s$.
- Let $c_{max}$ denote the maximum concentration allowed at the measure point. Because $c(x)$ depends linearly on $Q$ (see section 4) the allowed emission velocity $Q_a$ is:

$$Q_a = \frac{c_{max}}{c(x)} 100 \quad mg/s$$

# 6 Conclusions

The Gaussian model is well known throughout literature and it is also widely spread among those working on air pollution problems. The reasons for that are the following:
- it has an analytical solution
- it is a natural model, meaning that, given some simplifications - like the ones we made - you arrive exactly at this model.

The fact that it is the only model known so far to be of any use will also have had an effect, we believe.

The solutions to the model also give the solutions for our problem. We do however recommend to check these, because some possibly significant effects were supposed to be negligeable or not taken into account for simplification purposes. This means that parameters may have to be adjusted to the actual values.
For instance, chemical reactions were left out. Also it may be the case that our worst case analysis will never actually appear. We may therefore arrive at an overestimated maximum output when using the solution as it is now.
In our case, with an already existing plant, this means looking at the data collected from the past, if available, and adjust the parameters accordingly. If not available, data collection has to be the first step.

There are four classes of weathertypes defined in Chapter 4, one of those only occurring at night. You can see from the results in Chapter 5, that the different weathertypes have quite some influence on the maximum output. The company therefore could set up some 'active emission policy': when the weather is unstable the emission rates can be higher than with a stable weathertype. The policy then could be to adjust the production to the maximum output.

So after checking the solutions the company has the following options:
- reduce emission such that even in the case of stable weather the norm will not be exceeded
- reduce emission such that, given a certain weathertype, the norm will not be exceeded, and than adjust production to changes in types of weather.

# References

[Seinfeld 86] Seinfeld, John H.
Atmospheric Chemistry and Physics of Air Pollution,
John Wiley & Sons, New York, 1986.

[Zeedijk 91] Zeedijk, H.
Atmosfeer & Luchtverontreiniging, (in Dutch)
lecture notes University of Twente, ED1191, 1991.

# A VEHICLE ROUTING PROBLEM IN NOORD-BRABANT AND LIMBURG

# How to schedule trucks in a distribution problem

A supermarket has to provision branches in Noord-Brabant and Limburg. Trucks are rented to deliver the demands of the branches. Until now, the truck routings needed have been composed by hand. We were asked to make a computer program that produces better, i.e., cheaper, routings than those made by hand. This means that the program has to reduce the costs for hiring trucks. These costs consist of three parts: a fixed part per day, a variable part for the *time* a truck is used, and a variable part for the *distance* driven in it. An extra complication is the fact that the trucks have different capacities.

To find a good routing for this problem is difficult. In fact, it can be proven that there is no way to obtain an optimum solution fast. That is why we use heuristics to solve the problem. In this report we describe two methods, the first is called Clarke & Wright's method; the second is based on finding the Hamiltonian Cycle through all the delivery points.


- Clarke & Wright's method – turn to Section 4.2, page 12
- Hamiltonian Cycle based method – turn to Section 4.3, page 14


We also produced a routing by hand to see what the important factors for minimizing the costs are and to be able to compare it with the solutions found through the heuristics.

# Results of the routing programs and some recommendations on further research

The best heuristic was Clarke & Wright's method. However, differences are small: total costs were *f* 5638,- compared to *f* 5704,- for the other routings.

Some advantages of the computer implementation of the routing problem are:

- it produces cheaper routings,

- there is no need any more for a specially trained employee to produce the routings,
- integration of the routing program into a larger software package – for job scheduling of truck drivers for instance – is possible.

We can think of two ways to produce even better routings and we recommend further research in these directions. The first way is to develop a heuristic that starts from a solution as the one obtained with Clarke & Wright's method and then improves it. The second way is to develop a better algorithm: Clarke & Wright's method is a very basic algorithm.

Before doing this, we think it would be useful to obtain a lower bound on the total costs to see what improvement may be possible. If this lower bound is close to the costs already found, it may be that no further research is necessary.

# Contents

# 1 Distributing pallets in Noord-Brabant and Limburg, an outline of the problem

## 1.1 Introduction

A supermarket distribution centre prövisions branches in Noord-Brabant and Limburg several times a week. The branches place their orders two days in advance. The orders are then collected and put on pallets, ready for transport.

Trucks are rented from several transport firms for transporting the pallets. To make sure that trucks will be available, this should be done the afternoon before. So what we need to know is how many trucks we want to rent and of what type - there are trucks with a capacity of 20, 28 or 40 pallets, respectively. When you want to do this, you have to decide what type of truck is sent to which city. And it has to be done in an optimal way, meaning, at minimal costs. Costs come from renting a truck (fixed costs per day and variable costs per hour using the truck and per kilometre driven in it) and from unloading and loading pallets (time costs).

## 1.2 Purpose

Until now, the routings of the trucks have been composed by hand. We want to implement a computer program that produces cheaper routings. Because of the structure of the problem, we apply heuristics to find a solution.

First we formulated the problem mathematically, i.e., as a 0-1 linear programming problem. This is known in the literature as VRP (Vehicle Routing Problem). We also did literature studies on the problem. As a result of that we found a heuristic proposed by Clarke & Wright ([1]), which we implemented in a computer program. Further more we developed a method of our own, suggested by Sergey Tiourine, a fellow student. It is based on finding the Hamiltonian cycle of the cities the branches are in.

## 2 Figures provided by the distribution centre

The distribution centre provided us with the data we needed to compute a solution. The data consists of:

- costs for hiring the trucks
- average driving speeds
- time for loading and unloading
- demand of the branches on a specific day.

In Table 1 you can see we have a choice among three different types of trucks: trucks with a maximum capacity of 20, 28, or 40 pallets, respectively. Costs for hiring them consist of a fixed part per day and two variable parts: one for the time you use the truck and one for the number of kilometres you drive in it.

| Type of truck | Number of pallets | Payment in guilders | | |
|---|---|---|---|---|
| | | per day | per hr | per km |
| 1 | 20 | 160 | 50 | 0.45 |
| 2 | 28 | 200 | 50 | 0.50 |
| 3 | 40 | 240 | 50 | 0.55 |

Table 1: Costs of renting a truck (in guilders).

Average driving speeds are road dependant. On highways it is 80 kph, in cities it is 20 kph, and on other roads it is 60 kph on average. Here kph mean: kilometres per hour. We take them independent of driving directions, type of truck, and load of a truck.

For loading and unloading you need on average a fixed time of 10 minutes plus some time per pallet. This extra time is 1 minute per pallet when loading and two minutes per pallet when unloading.

The one-day distribution list we obtained from the distribution centre is as follows:

| Number of pallets per branch | | |
|---|---|---|
| 1 | Bergen op Zoom | 10 |
| 2 | Boxtel | 7 |
| 3 | Breda | 20 |
| 4 | Den Bosch | 18 |
| 5 | Deurne | 12 |
| 6 | Dongen | 10 |
| 7 | Echt | 9 |
| 8 | Eindhoven | 21 |
| 9 | Geleen | 14 |
| 10 | Heerlen | 19 |
| 11 | Helmond | 12 |
| 12 | Maastricht | 18 |
| 13 | Oosterhout | 15 |
| 14 | Oss | 17 |
| 15 | Roermond | 14 |
| 16 | Roosendaal | 11 |
| 17 | Sittard | 18 |
| 18 | Tilburg | 19 |
| 19 | Uden | 14 |
| 20 | Valkenswaard | 11 |
| 21 | Veghel | 9 |
| 22 | Venlo | 16 |
| 23 | Venray | 13 |
| 24 | Waalwijk | 15 |
| 25 | Weert | 14 |

Table 2: A one-day ordering list.

Finally, we needed a distance table, and a classification of the roads in highway and normal road. We set the distances driven within cities to 2 km (so in and out takes 4 km driving), except for Breda, Eindhoven, Heerlen, Maastricht, Tilburg, and Venlo, where we put them to 4 km. The other distances came from a road map of the Netherlands, from which we also estimated that roughly 80 % of the roads were highways. So if for instance we have to drive 55 km with a truck, we take that to be $\frac{1}{5} \times 55 = 11$ km normal road and $\frac{4}{5} \times 55 = 44$ km highway.

# 3 Mathematical formulation of the Vehicle Routing Problem (VRP) resulting from the distribution problem

We adapted the following formulation of the vehicle routing problem (VRP) from [2]. It uses binary variables to indicate whether a truck travels between two given cities in the optimal solution. We numbered the cities according to Table 2 in Section 2 (page 6). City 0 is Maarheeze, the city where the depot is.

For the mathematical formulation we introduce the following variables:

$m$ : number of available trucks, (in our case it can be infinity)

$k_1$ : number of trucks of type 1,

$k_2$ : $(k_2 - k_1)$ trucks of type 2,

$(m - k_2)$ trucks of type 3,

$D_k$ : capacity of truck k ($D_k$= 20, 28 or 40), (k = 1,..,m)

$f_k$ : fixed costs of truck k ($f_k$= 160, 200 or 240), (k = 1,..,m)

$d_i$ : number of pallets ordered in city i (i = 1,..,25)

$t_{ij}$ : time needed to travel from i to j in hours (i,j = 0,..,25)

$(t_{ii} = \infty)$

$\epsilon_i$ : distances driven within city i in kilometres

$y_{ij}$ : distance between i and j in kilometres (i,j=0,..,25)

$L$ : time allowed for a route in hours (in our case L = 8)

$$x_{ij}^k = \begin{cases} 1 & \text{if truck k travels directly from i to j} \\ 0 & \text{otherwise} \end{cases} \quad \text{(i,j=0,..,25; k=1,..,m)}$$

Now we can formulate our VRP as follows:

objective function:

$$\text{minimize} \quad \sum_{k=1}^{m} f_k \sum_{j=1}^{25} x_{0j}^k + 50 * \sum_{i=0}^{25}\sum_{j=0}^{25}\sum_{k=1}^{m} t_{ij} x_{ij}^k$$

$$+ \quad \frac{50}{60} * \left( \sum_{i=0}^{25}\sum_{j=1}^{25}\sum_{k=1}^{m}(10 + 2 * d_j)x_{ij}^k + \sum_{k=1}^{m}\sum_{j=1}^{25} 10 x_{0j}^k + \sum_{j=1}^{25} d_j \right)$$

$$+ \quad 0.45 * \sum_{i=0}^{25}\sum_{j=0}^{25}\sum_{k=1}^{k_1}(y_{ij} + \epsilon_j)x_{ij}^k$$

$$+ \quad 0.50 * \sum_{i=0}^{25}\sum_{j=0}^{25}\sum_{k=k_1+1}^{k_2}(y_{ij} + \epsilon_j)x_{ij}^k$$

$$+ \quad 0.55 * \sum_{i=0}^{25}\sum_{j=0}^{25}\sum_{k=k_2+1}^{m}(y_{ij} + \epsilon_j)x_{ij}^k$$

The first term in the objective function specifies the fixed costs, the second term describes the costs for the total time the trucks are used. The next six terms are the costs for the driven kilometres, where the first three specify the costs for driving from one city to another and the last three for driving inside the cities. The remaining terms describe the costs for loading and unloading the trucks.

constraints:

$$\sum_{i=0}^{25}\sum_{k=1}^{m} x_{ij}^k = 1 \qquad (j = 0,..,25;\ k = 1,..,m) \qquad (1)$$

$$\sum_{i=0}^{25} x_{il}^k - \sum_{j=0}^{25} x_{lj}^k = 0 \qquad (k = 1,..,m;\ l = 0,..,25) \qquad (2)$$

$$\sum_{i=0}^{25}\sum_{j=1}^{25} d_j x_{ij}^k \leq D_k \qquad (k = 1,..,m) \qquad (3)$$

$$\sum_{i=0}^{25}\sum_{j=0}^{25} t_{ij}\frac{x_{ij}^k}{60} + \sum_{i=0}^{25}\sum_{j=1}^{25}(10 + 3z_{jk})\frac{x_{ij}^k}{60} + \sum_{j=1}^{25} 10\frac{x_{0j}^k}{60} \leq L$$

$$(k = 1, .., m) \qquad (4)$$

$$\sum_{i=1}^{25} x_{i,0}^{k} \leq 1 \qquad (k = 1, .., m) \qquad (5)$$

$$\sum_{j=1}^{25} x_{0,j}^{k} \leq 1 \qquad (k = 1, .., m) \qquad (6)$$

$$\sum_{i \in S} \sum_{j \in S} x_{ij}^{k} \leq |S| - 1 \qquad (\text{for all } S \text{ s.t. } |S| \geq 1; \quad S \subseteq \{1, .., 25\}; k = 1, .., m)$$

$$x_{ij}^{k} = 0, 1 \qquad (i, j = 0, .., 25; \quad k = 1, .., m) \qquad (7)$$

The first and second constraint specify that each city must be served exactly once by one and the same vehicle. The third constraint guarantees that truck capacities are never exceeded; constraint number four ensures that no truck route exceeds its time limit. The fifth and sixth constraint ensure that no more than m trucks leave the depot. The seventh constraint specifies that a truck routing has to start and end in Maarheeze and has to consist of one piece. The last constraint specifies that it is a 0-1 formulation.

# 4 Three methods for solving the VRP problem

In this paragraph we describe three methods for solving the VRP. First we present the handmade solution. Then we give a method described by Clarke & Wright, which we found in literature and finally we show a method we developed ourselves. We tested all three methods and we will give some results in the next section.

## 4.1 Obtaining truck routings by hand

To get a solution by hand, we thought we ought to start with trucks of largest capacity and send them 'far away'. We then searched in the neighbourhood for cities that still need pallets and include them in the route. We did it in such a way that we completely filled the trucks, with as less splitting as possible (when more than one truck serve a city, we call that splitting).

Doing this we got the solution shown in Table 3. The numbers between ( ) denote the number of pallets delivered by a truck to the city number in front of it. **Total pallets** denotes the load of a truck. For instance, route number two consists of driving a truck loaded with 40 pallets from Maarheeze first to Geleen, then from Geleen to Echt, from Echt to Sittard, Sittard to Valkenswaard, and then driving back to Maarheeze (or vice versa).

As can be seen in the table there are 7 routes with load 40, 2 with load 28, and 1 with load 20. This would mean that we need 7 trucks of capacity 40, 2 of 28, and 1 of 20, unless we can use one truck for two or more routes. I.e., two or more routes can be executed within eight hours by the same truck. Also you can see that two cities were split: Geleen is in route no 1 as well as in route no 2, and Valkenswaard is both in route no 2 and route no 7.

We will not discuss combining two routes here, but in Numerical Results and Comparisons (see Section 5, page 16), where we also present the costs for this routing schedule.

| route | city | (# pallets) | time(min) | total pallets |
|---|---|---|---|---|
| 1 | Maastricht | (18) | | |
| | Heerlen | (19) | | |
| | Geleen | (3) | 389 | 40 |
| 2 | Geleen | (11) | | |
| | Echt | (9) | | |
| | Sittard | (18) | | |
| | Valkenswaard | (2) | 398 | 40 |
| 3 | Bergen op Zoom | (10) | | |
| | Roosendaal | (11) | | |
| | Tilburg | (19) | 433 | 40 |
| 4 | Waalwijk | (15) | | |
| | Oosterhout | (15) | | |
| | Dongen | (10) | 359 | 40 |
| 5 | Uden | (14) | | |
| | Oss | (17) | | |
| | Veghel | (9) | 322 | 40 |
| 6 | Helmond | (12) | | |
| | Deurne | (12) | | |
| | Venlo | (16) | 351 | 40 |
| 7 | Den Bosch | (18) | | |
| | Valkenswaard | (9) | | |
| | Venray | (13) | 393 | 40 |
| 8 | Weert | (14) | | |
| | Roermond | (14) | 212 | 28 |
| 9 | Boxtel | (7) | | |
| | Eindhoven | (21) | 237 | 28 |
| 10 | Breda | (20) | 238 | 20 |

Table 3: Routings found by hand

## 4.2 Clarke and Wright's method

The first method we found was the one formulated in an article from Clarke and Wright [1], so we named it Clarke & Wright's method.

We will only give the formulation of the method, not a discussion of why it is as it is. For this we recommend reading the article mentioned above, which is very clear and well-readable.

### 4.2.1 Formulation

It is the VRP problem again, so there are 25 trucks available, with capacity $D_k$ $(k = 1, .., n)$ and demand $d_j$ required to be delivered to city j $(j = 1, .., 25)$ from a depot (city 0).

Given the distances $y_{ij}$ between all the cities we want to minimize the total distance covered by the trucks. In doing so, we also minimize driving times and thus the total costs. This is because Clarke & Wright's method does not allow splitting, and so all costs besides driving costs are fixed.

In the article the assumption is made that $D_{max} \geq d_j(\forall j)$ which of course is no restriction to the problem: if this is not the case simply send a truck with the largest capacity to such a city until this assumption is valid. Also it is assumed that $D_{max} \ll \sum_{j=1}^{25} d_j$. Both assumptions are valid in our case.

### 4.2.2 Implementation

The idea of the method is to assign a truck to each city first and then combine two routes in such a way that a maximum reduction of costs is realized in such a step. It can be proven that the best way to combine two routes is when you go from the end of one route to the start of the next one (or other combinations of starting and endpoints). The reduction of costs when combining two routes with cities $i$ and $j$ then becomes: $y_{0i} + y_{0j} - y_{ij}$ (denoted by $s_{ij}$, the so called shadowcosts).

To see if a truck has to drive from $i$ to $j$ we introduce the variable $t_{ij}$, which is 1 in that case, and 0 if the two cities are not linked in the route of a truck.

When a city $i$ is exclusively served by one truck, $t_{0i} = 2$. The initial solution therefore will be: $t_{0i} = 2(i = 1, .., 25)$.

Now we proceed as follows:

1. Find the maximum shadowcost;
2. If the combined load is smaller or equal than the maximum capacity one can use then combine the two routes (adjust $t_{ij}$'s and load of the truck) and adjust the shadowcosts;
3. If it is bigger, set the shadowcost found to zero;
4. Repeat steps 1 to 3 until all the shadowcosts are zero.

From the implementation it is clear that the algorithm prefers using large trucks. To see what happens when you use less of those, it is possible to put restrictions on used number of certain truck types. See also Numerical Results and Comparisons (page 16).

## 4.3 Method based on Hamiltonian Cycle

### 4.3.1 Introduction

The main idea of the method is as follows:

First compute the shortest Hamiltonian cycle. The Hamiltonian Cycle is a cycle that connects all the cities in such a way that the cycle has minimal total distance. To obtain the Hamiltonian the central depot in Maarheeze is omitted. Then we divide the cycle into parts such that a truck can deliver to all the cities in one part.

Now a route for a truck is the following: from Maarheeze to the starting point of a part, the part itself and from the end point of the part back to Maarheeze.

### 4.3.2 Implementation

To find the shortest Hamiltonian cycle we use a heuristic defined in [3]. This algorithm solves the Traveling Salesman Problem. There are many possibilities for dividing the cycle in parts when we allow splitting cities. Therefore we use a local optimization algorithm:

For the location of the first starting point we choose a city (25 possibilities) and a direction on the cycle (2 possibilities). This will give us 50 solutions.

Now we have to decide what truck is used in this part and wether a city is split to which the complete demand can not be delivered or not. The local search algorithm we defined does this in the following way:

For each type of truck the complete demand of a city is delivered to as many cities as possible. Now there are two possibilities:

- Fill the truck completely. This means splitting of the next city.
- Do not go to the next city. This means you do not use the whole truck capacity.

To decide which of the two options we take, we define a parameter $\alpha$ such that the minimal load of a truck is greater or equal than the truck capacity minus $\alpha$.

Furthermore we made some rules that say which kind of truck we preferred. For instance, we think it is better to fill a truck completely than have some empty space, and that a completely filled truck with capacity 40 is relatively cheaper than a completely filled one of 28. Applying these rules you get a list of possibilities:

1. a truck with capacity 20 serving all the remaining cities;
2. a truck with capacity 28 serving all the remaining cities;
3. a truck with capacity 40 serving all the remaining cities;
4. a truck with capacity 40 delivering all cities in the part completely ($\leq \alpha$ places open);
5. the same for a truck with capacity 28;
6. the same for a truck with capacity 20;
7. a truck with capacity 40 delivering to the last city in the part at least half of the demand and is completely filled;
8. the same for a truck with capacity 28;
9. the same for a truck with capacity 20;
10. a truck with capacity 40 delivering to the last city in the part less than half of the demand (still completely filled);
11. the same for a truck with capacity 28;
12. the same for a truck with capacity 20.

Each iteration we prefer the type of truck that has the lowest number in this list. In the implementation we do this with the help of a weight function. After the truck for this part is chosen, we start the same procedure again with the remaining cities and demands, until all demand has been assigned to a truck.

# 5 Numerical Results and Comparisons

To compare the solutions, we wrote a procedure to calculate the total costs for every solution. The program also calculates the total time and distance for every route in that solution. Here we use the assumption that 80 % of the road between every city (given in the distance table) is highway.

Before calculating the costs, however, we wrote a computer program that checks if routes can be 'combined', i.e., can be done within eight hours using one truck. Combining two truck routes is always cheaper because less trucks have to be rented. Therefore it is also better to combine two trucks of different capacity and use the one truck with biggest capacity.

In case of the solution we found by hand, for instance, we found that route no 8 and route no 9 can be done by one truck (see Table 3). Total time then is $7\frac{1}{2}$ hours (449 minutes to be exact). So the best solution we could find by hand is:

| hand | | | total costs = 5704 | | | | | |
|---|---|---|---|---|---|---|---|---|
| nr | type | load | route | | | | time(min) | km |
| 1 | 40 | 40 | 12(18) | 10(19) | 9( 3) | | 389 | 186 |
| 2 | 40 | 40 | 9(11) | 7( 9) | 17(18) | 20( 2) | 398 | 187 |
| 3 | 40 | 40 | 1(10) | 16(11) | 18(19) | | 433 | 256 |
| 4 | 40 | 40 | 24(15) | 13(15) | 6(10) | | 359 | 179 |
| 5 | 40 | 40 | 19(14) | 14(17) | 21( 9) | | 322 | 133 |
| 6 | 40 | 40 | 11(12) | 5(12) | 22(16) | | 351 | 154 |
| 7 | 40 | 40 | 4(18) | 20( 9) | 23(13) | | 393 | 221 |
| 8 | 28 | 28 | 25(14) | 15(14) | | | | |
| | | | 2( 7) | 8(21) | | | 449 | 176 |
| 9 | 20 | 20 | 3(20) | | | | 238 | 168 |

Table 4 : Best solution for routings found by hand.

The numbers in route denote the city number as given in Table 2 on page 6 together with the number of pallets delivered to that city between ( ). This table can be used as driving schedule for the truck drivers.

Running the 'combination' program for every solution found by all methods we obtain the following list of 'best' solutions.

| Method | Truck type | Costs |
|---|---|---|
| Hand | (1,1,7) | 5704 |
| Clarke & Wright | (0,1,8) | 5638 |
| Hamiltonian Cycle ($\alpha = 0$) | (0,2,7) | 5918 |
| Hamiltonian Cycle ($\alpha = 1$) | (1,4,4) | 5901 |
| Hamiltonian Cycle ($\alpha = 2$) | (2,2,5) | 5772 |
| Hamiltonian Cycle ($\alpha = 3$) | (1,1,7) | 5704 |
| Hamiltonian Cycle ($\alpha = 4$) | (1,1,7) | 5704 |
| Hamiltonian Cycle ($\alpha = 5$) | (1,2,6) | 5723 |

Table 5: Total costs of solutions of the several methods.

The vector notation for the truck type means for example in the Clarke & Wright solutions that we use no trucks with capacity 20 and that we hire one truck with capacity 28 and eight trucks with capacity 40.

Comparing the solutions we see that the method of Clarke & Wright gives the best result. The solution found by hand and the best solution of the method based on the Hamiltonian Cycle are exactly the same. But the differences between the costs of the solutions found by the three methods are very small.

# 6 Conclusions and Recommendations

The problem of finding truck routes from a central depot to a number of delivery points has been solved by two heuristic methods. For the one example we got the method of Clarke & Wright is the best. But the differences with the method based on the Hamiltonian Cycle and the solution found by hand are relatively very small. So it is hard to say which method is the best.

If the number of delivery points would increase in the future the two implemented methods will still work correctly while it will become difficult to obtain good solutions by hand. Another advantage of the implementation is that it is easy to print the routes for the truck drivers together with the time they need for making a route.

The methods can be improved in a few ways to obtain better solutions. However, we think it is better to compute a lower bound for the costs. This lower bound then indicates the need and/or possibility for improvement of the methods.

A lower bound can be obtained by solving the VRP-problem by means of an integer linear programming package.

An improvement for the method based on the Hamiltonian Cycle could be the following one. Each time one truck route has been defined one repeats the procedure for obtaining the Hamiltonian Cycle without the just served cities. In this way the costs for the number of kilometres driven decreases. But we do not know if this decrease will influence the present solution very much. For sure it will take a lot of more computation time.

The only thing we can think of for improving the method of Clarke & Wright is to allow 'splitting' cities. But in literature we could not find anything that handles splitting so probably it is not possible for this method.

# References

[1] G. Clarke and J.W. Wright
Scheduling of vehicles from a central depot to a number of delivery points
Operations Research, vol. 12, pp. 568-581, 1964

[2] G. Laporte and Y. Nobert
Exact algorithms for the vehicle routing problem
Annals of Discrete Mathematics, vol. 31, pp. 147-184, 1987

[3] Lecture notes "Optimaliseringsalgorithmen" (158009), University of Twente, 1990

# SERVING THE SCIENCE MUSEUM

# Summary

The National Science Museum makes science and technology accessible to ordinary people. The museum welcomes many groups every day, each of which either has a professional guide or, at busy times, is guided by a part-time hired employee or a member of the scientific staff. Although regular guides are preferred and cheaper than the instant forces, nowadays part-timers accompany half the groups. So the directorate wants to know how to schedule professional guides regarding the number of groups while reducing costs and keeping the work of guides acceptable.

According to us, the museum has two directions in which to improve the guiding achievements. Firstly they can utilise more guides than the 8 working there now. Next, the working schedule for guides of 7 consecutive working days in a period of 10 days can be altered. With two variables, the number of guides and the working schedule, an optimum with respect to costs is found. A schedule with a period of 7 days and 5 working days is always preferable to the old schedule. Then the museum can best employ 12 guides.

The next aspect is to derive a working scheme for the guides restricted to social aspects. The aspects are:

- the maximum number of consecutive working days
- the number of weekends off
- the ease of reading a scheme for the whole year.

For the optimal schedule with 12 guides each guide starts working on a specific day, and he shifts the next week to another starting day.

# Table of contents

# 1 The National Science Museum

## 1.1 Welcoming groups

The task of the National Science Museum is to bring science and technology closer to the civilians. A special goal is to interest students for technical studies.

The Science Museum welcomes many groups every day. Every group is guided by one guide during the whole day. Therefore the museum employs a number of regular guides who are skilled in accompanying these groups. The guides are present when their working schedule indicates they should be, so the number of guides normally is not adjusted for the number of groups arriving. If the last quantity exceeds the number of regular forces present, the museum approaches part-time servants. These servants may or may not be available. If enough guides still are not available, members of the normal scientific staff can help out. At the moment, an arriving group is promised to be guided if they give notice two days in advance.

Before reporting the current problem, looking at some major figures is useful. The numbers give insight in the relevance of the points mentioned above and enlighten the understanding of the serving problem.

## 1.2 The museum data

As the task of the museum is stated above, we can look for the information about the way the museum has performed until now. Therefore the directorate gave us figures about the operating the last 8 weeks. These data deal with the scheduling of guides, the arrival of groups and the payments and absence of the personnel.

- First, the regular guides have been scheduled according to a fixed schedule of 7 days on duty and 3 days off duty.
- In the past 8 weeks, the numbers of groups per day have been counted. These numbers are shown in Table 1.1. The list of figures shows the scholars are not predictable in visiting the museum. Although in general, they have preference to visit Wednesdays and Fridays.

Table 1.1 Group arrivals in past 8 weeks

| MON | TUE | WED | THU | FRI | SAT | SUN |
|---|---|---|---|---|---|---|
| 6 | 9 | 12 | 11 | 12 | 9 | 5 |
| 7 | 7 | 13 | 10 | 13 | 8 | 8 |
| 9 | 10 | 14 | 11 | 15 | 7 | 7 |
| 5 | 8 | 11 | 9 | 17 | 10 | 11 |
| 4 | 9 | 15 | 10 | 10 | 8 | 4 |
| 7 | 6 | 14 | 7 | 15 | 11 | 8 |
| 8 | 11 | 16 | 12 | 9 | 9 | 7 |
| 7 | 8 | 13 | 9 | 14 | 5 | 9 |

- Also the number of times that part-timers have been approached in that period is written down: 426 part-timers were phoned up, 315 (74%) of whom could be reached and 226 (53%) of them were available to guide.
- The regular guides receive *f*50,000.= a year. The part-timers are paid *f*300.= a day and the scientific staff are paid *f*80,000.= a year.
- Sick leave is approximately 4% and the guides are obliged to take their holidays within school holidays.

## 1.3 Scheduling guides

The serving of scholar groups is done by three types of employees, regulated by succeedingly taking employees from some reservoirs. The order of selecting types to serve is not arbitrary. Part-timers, mostly students, are less professional in the guiding task than regular guides. The scientific forces may be competent in guiding groups but are employed for another task in another department of the museum. Next to these aspects, both the part-timers and scientific staff are more expensive than the guides who are used to lead groups around. You can compare the costs of the different employees by taking into account a working period of about 200 days a year. Then the price for a normal guide is about *f*250.= a day, and for a scientist more than *f*400.= a day. Consequently the directorate prefers regular guides above part-timers, and using part-timers above hiring scientific employees.

Despite the preference of serving groups by regular guides, frequently half of the groups of scholars are served by part-timers. The directorate considers an increment of the number of the regular guides to be useful, both for costs and serving aspects. This conviction is basis for the request to investigate the possibilities of improving the employment, with regard restricted to the costs of personnel. The suggested improvement is an increase of the number of regular guides. Though maybe other variables can be altered to reduce costs. Next to the aim of reducing costs, the directorate pays attention to the social aspects. A new system for scheduling guides should be socially acceptable for the guides themselves. Therefore, the suggestions for improvement should include an observation of these aspects.

## 1.4 Orienting approach

The modelling of this problem consists of several steps. Before modelling all aspects, we orient on the present situation with a simple model. The group arrival is considered to be deterministic in the next chapter. That simplification provides an easy way to look at the important parameters for lowering costs. Then, a more sophisticated model for the group arrival helps to find the optimal parameters.

In the first two steps no attention is paid to social aspects. Until in chapter 4, this aspect is included to find an optimum. The optima found without social restrictions or social goals, are elaborated to be socially acceptable.

# 2 A simple, deterministic model

## 2.1 Further assumptions

The problem is to find a more cost efficient organisation concerning the guides. The first step is to analyse the organisation in the present situation. Therefore we made the following assumptions:

- Every group reserves, so every group gets a guide.
- Subtracting the number of school holidays from the number of days in a year leads to 280 working days in a year (weekends included).
- The regular guides work in a period 280 days a year. Due to the schedule of 7 days work - 3 days off, the yearly payment corresponds with $f50{,}000/(280\,(7/10)) = f255.10$ a day.
- The scientific staff work the same number of days as the regular guides; $280\times(7/10)$ days in a year. We assume they get paid by the guiding department of the museum for every day they help to guide, which means they get paid $f80{,}000/(280\times(7/10)) = f408.16$ every day. We also assume that there is enough scientific staff such that every arriving group can get a guide.
- After a school holiday, a schedule proceeds as if nothing happened.

It is given that the regular staff works for 7 days and then is 3 days off duty. We call these 7+3 consecutive days a **period**. When a guide starts with a period on the first January, he can compute exactly on which days of the year to work and which not. This way we can define a **schedule** to be an array of ten numbers, where the $i^{th}$ number is equal to the number of guides that start their period on the $i^{th}$ day of the year. For example:

0 0 2 0 1 1 1 0 1 2

One can prove that when there are 8 regular guides,

$$\frac{1}{7+3}\binom{7+3-1+8}{8} = 2431$$

schedules are possible. This is a rather large number of schedules. So even in this deterministic case, the computation time will be extensive.

To reduce computation time, we make a first approximation of the present situation by excluding all stochastics from the problem. In some problems, results derived from the deterministic case are as good as the results in the stochastic case. We do not have any indication of the righteousness now, but we certainly can obtain a better view of the important parameters in the problem than we have right now.

We therefore make the following simplifying assumptions:

- Assume the number of arriving groups on a day is deterministic: On every day the number of groups is given by the mean of the outcomes from Table 1.1 (previous section).
- Assume that there is no sick leave.
- On every day 15×53% = 8 part-timers are available.

## 2.2 Guide schedules

We define the following constants and variables:

$n$ is defined to be the length of a working period for each regular guide.
$k$ is the number of consecutive days on duty for a regular guide.
$f$ the number of regular forces the museum employs.

Given a schedule, we can compute the number of working regular guides on a day. We assumed the number of groups that will arrive on each day of a week to be deterministic and known. When the number of regular staff is not sufficient to guide each group, some of the part-timers have to work, and when even this number is not sufficient, the rest of the groups will have to be guided by a member of the scientific staff.

In this way, it is possible to compute the total costs in a year for each schedule.

We wrote a computer program that computes for every possible schedule the total costs in a year. The five best schedules (those schedules that give the lowest cost) will be found in the next section.

## 2.3 Analysis of the present situation

In the first running of the computer program, the constants have the following values:

$n = 10$ (length of a period)
$k = 7$ (number of consecutive days on duty for a regular guide)
$f = 8$ (number of regular forces)

The five best schedules appear to be:

| Schedule | | | | | | | | | | Costs: |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 2 | 0 | 1 | 1 | 1 | 0 | 1 | 2 | *f*739,438.37 |
| 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 2 | *f*739,438.37 |
| 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 2 | *f*739,438.37 |
| 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 2 | *f*739,438.37 |
| 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | *f*739,438.37 |

(In fact, about 100 more schedules led to the same optimal costs.)

## 2.4 Possible improvements

Now that we have an idea of the total expected costs in a year in the present situation, we can look for ways to improve this situation. Here we can think of:

- Changing the number of regular guides
- Changing the period of a "regular-guide-week"
- Changing the number of part-timers.

With respect to the third possibility:
Increasing the number of part-timers is always optimal: You only pay them when you really need them and the probability to have to hire scientific staff (who are more expensive) is decreased, so this will definitely reduce the costs. However, in the original problem description, this possibility was not suggested so we leave it.

The first and second possibility have been checked using the program. Some of the results can be found in the next section.

# 2.5 Analysis of the suggested improvements

Costs can be reduced by choosing different constant values, for example by increasing the number of regular guides. Why should increasing $f$ reduce costs? Therefore, imagine two extreme situations. First when there are no guides, you have to hire part-timers for all guiding tasks. As they are more expensive than guides this is far from optimal. Take on the contrary about 20 guides so that you never need a part-timer. Because the guides remain idle many times, this also is far from optimal. Thus the optimal number of guides is in between those extrema, like in a parabola. As in the present situation many part-timers are hired, we look for increase of $f$.

$n = 10$ (length of a period)
$k = 7$ (number of consecutive days on duty for a regular guide)
$f = 9$ (number of regular forces)

The five best schedules appear to be:

| Schedule | | | | | | | | | | Costs: |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 0 | 2 | *f*731,040.= |
| 1 | 0 | 2 | 1 | 0 | 1 | 2 | 0 | 0 | 2 | *f*731,040.= |
| 0 | 1 | 1 | 1 | 0 | 2 | 1 | 0 | 2 | 1 | *f*731,040.= |
| 0 | 1 | 2 | 0 | 0 | 2 | 1 | 0 | 2 | 1 | *f*731,040.= |
| 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | *f*731,040.= |

$n = 10$ (length of a period)
$k = 7$ (number of consecutive days on duty for a regular guide)
$f = 10$ (number of regular forces)

The five best schedules appear to be:

| Schedule | | | | | | | | | | Costs: |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | *f*725,600.= |
| 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | *f*727,040.= |
| 0 | 1 | 2 | 0 | 1 | 1 | 1 | 1 | 1 | 2 | *f*727,040.= |
| 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 1 | 2 | *f*727,040.= |
| 0 | 1 | 1 | 1 | 1 | 1 | 2 | 0 | 1 | 2 | *f*727,040.= |

For $f = 11$, costs increases again.

One can see that the best schedule, given $n = 10$ and $k = 7$, appears to be the one where the starting points of the periods of the guides are distributed equally among the first $n$ days of the year. This is intuitively clear because the cycle length of the arrivals of groups is equal to 7 days. The period length of the guides is 10 so these two numbers are coprime. This means that the high peaks of arriving groups can not be covered by high peaks of regular guides on duty, because these peaks shift with respect to each other. The best schedule is thus the schedule where the number of regular guides on duty is equal every day.

We can also conclude from the last section that it may be optimal to set $n$, the period length of the guides, equal to 7, which equals the cycle length of the numbers of arriving groups. In this situation, we can try to find a schedule such that the peaks in the number of regular guides do coincide with the peaks in the number of arriving groups. We can run the program with constants $n = 7$ and $k = 5$. This period equals a normal working week. The number of working days in a year, with these numbers, equals to $280 \quad 5/7 = 200$, and is was equal to $280 \quad 7/10 = 196$, so the forces have to work an extra 4 days.

$n = 7$ (length of a period)
$k = 5$ (number of consecutive days on duty for a regular guide)
$f = 8$ (number of regular forces)

The five best schedules appear to be:

| Schedule | | | | | | | Costs: |
|---|---|---|---|---|---|---|---|
| 3 | 4 | 0 | 0 | 1 | 0 | 0 | *f*728,800.= |
| 4 | 3 | 0 | 0 | 1 | 0 | 0 | *f*728,800.= |
| 5 | 2 | 0 | 0 | 1 | 0 | 0 | *f*728,800.= |
| 0 | 1 | 5 | 2 | 0 | 0 | 0 | *f*728,800.= |
| 1 | 0 | 5 | 2 | 0 | 0 | 0 | *f*728,800.= |

$n = 7$ (length of a period)
$k = 5$ (number of consecutive days on duty for a regular guide)
$f = 10$ (number of regular forces)

The five best schedules appear to be:

| Schedule | | | | | | | Costs: |
|---|---|---|---|---|---|---|---|
| 4 | 2 | 2 | 1 | 0 | 1 | 0 | *f*708,800.= |
| 2 | 5 | 1 | 1 | 0 | 1 | 0 | *f*708,800.= |
| 3 | 4 | 1 | 1 | 0 | 1 | 0 | *f*708,800.= |
| 4 | 3 | 1 | 1 | 0 | 1 | 0 | *f*708,800.= |
| 2 | 1 | 6 | 0 | 0 | 1 | 0 | *f*708,800.= |

$n = 7$ (length of a period)
$k = 5$ (number of consecutive days on duty for a regular guide)
$f = 12$ (number of regular forces)

The five best schedules appear to be:

| Schedule | | | | | | | Costs: |
|---|---|---|---|---|---|---|---|
| 3 | 2 | 4 | 0 | 0 | 2 | 1 | *f*690,000.= |
| 3 | 2 | 4 | 0 | 1 | 1 | 1 | *f*690,000.= |
| 3 | 2 | 4 | 0 | 2 | 0 | 1 | *f*690,000.= |
| 4 | 1 | 5 | 0 | 0 | 2 | 0 | *f*690,000.= |
| 4 | 2 | 4 | 0 | 0 | 2 | 0 | *f*690,000.= |

These last results also give the optimal result, given $n = 7$ and $k = 5$. So, the optimal number of guides in a 7 day period is 12. For the initial period length 10, the optimal team contained 10 guides. The higher optimum is explained by the reason for altering $n$ and $k$. The argument for altering the working period was the fact the period didn't match the cycle length of the group arrival. Taking $n = 7$ it does match, and therefore the number of guides can increase to cover the peaks in the arrivals.

A result of the covering is the reduction of costs. Although, the decrease has to be adjusted for the change in total working days. With *(n,k) = (7,5)* the guides work 4 days per year longer and should be paid $4 \times f255.10 = f1020.40$ per guide more. The correction for 12 guides results in *f*12,244.80. Then the decrease in costs from *f*725,600.= to *f*690.000 is more than the adjustment, so the altering of the working period will be useful.

# 3 A more realistic, stochastic model

## 3.1 Motivation

Up to now, all stochastics were removed from the problem. To see what the impact is of this assumption on the results, we can derive a measure for the variances of the arrival processes. If we know that these variances are rather large, it may be useful to consider a non-deterministic model also.

To do this, we let the computer program, used in the previous sections, run again. However, instead of taking the mean value of the numbers of arriving groups every day, we take the minimal value, as observed in the last 8 weeks (see Table 1.1). We took $n = 7$, $k = 5$ and computed the optimal value to $f$:

$n = 7$ (length of a period)
$k = 5$ (number of consecutive days on duty for a regular guide)
$f = 12$ (number of regular forces)

The five best schedules appear to be:

| Schedule | | | | | | | Costs: |
|---|---|---|---|---|---|---|---|
| 3 | 2 | 5 | 0 | 1 | 1 | 0 | f600,000.= |
| 4 | 1 | 5 | 0 | 1 | 1 | 0 | f600,000.= |
| 5 | 0 | 5 | 0 | 1 | 1 | 0 | f600,000.= |
| 2 | 4 | 4 | 0 | 1 | 1 | 0 | f600,000.= |
| 3 | 3 | 4 | 0 | 1 | 1 | 0 | f600,000.= |

As one can see, the costs differ considerably from the comparable costs in the previous section (they were f690,000). We conclude that it is useful to look at the stochastic case, and will do this in the following paragraphs.

# 3.2 Group arrivals

Since in this chapter we drop the assumption that the number of groups arriving in a day is deterministic, we have to attain a suitable probability distribution to describe this number. This can be done in two ways:

- We can try to fit a regular probability distribution, such as a Gaussian or a Binomial one, by trying to find suitable parameters.
- We can define the table of observations of the number of arriving groups each day given in the first chapter "to be" the probability distribution.

First, we have to develop the first option before choosing. We have to find the best distribution describing the arrival of groups, so that it is possible to choose either the distribution or the fitting with the observations.

For choosing the most likely distribution see the figures of Table 1.1. For each day of the week, the interval in which the data appear differs significantly. The number of groups seems to depend upon the day of the week. That means we need a certain probability function for each day. These functions are difficult to derive, since we only have eight outcomes. Nevertheless, we can choose functions from two sorts of distributions: discrete or continuous. Because arrivals only take integer values, a continuous function has to be discretised. Since we have to estimate a function with eight data, an estimation plus discretisation will imply much work for a rather rough fitting. A continuous function has even more disadvantages: negative outcomes are not allowed. So, we choose a discrete function.

A test for the fitting will not be reliable since there is insufficient data. So we select the type by observation, and looking at the figures, the arrival seems likely to have a Binomial distribution. We need a distribution for each day, and therefore have to estimate 7 pairs $(n,p)$. The parameter pairs are based on the following unbiased estimates:

$$\hat{p} = 1 - \frac{s_x}{\bar{x}}, \qquad \hat{n} = \frac{\bar{x}}{\hat{p}} \tag{2}$$

Since $n$ has to be an integer we take the two nearest integers $\hat{n}_1 = \lfloor \hat{n} \rfloor$ and $\hat{n}_2 = \lfloor \hat{n} \rfloor$, and adjust $p_1$ and $p_2$ for the roundings by the relation $\hat{p}_i = \frac{\hat{n}_i}{\bar{x}}$. The last action needed to obtain 7 distributions is to choose between $(n_1,p_1)$ and $(n_2,p_2)$. Although we could not use statistical tests for the goodness of fit for distributions, we use it here. We utilise the Chi-Square Goodness-of-Fit Test as criterion for the best fitting Binomial parameters. The explanation of

this test and the results will be found in the appendix. The pairs *(n,p)* found for the 7 weekdays will be provided here.

**Table 1.2** Parameters for the Binomial distribution of group arrivals

| Day | n | p |
|---|---|---|
| MON | 9 | 0.736 |
| TUE | 11 | 0.773 |
| WED | 16 | 0.844 |
| THU | 12 | 0.823 |
| FRI | 16 | 0.820 |
| SAT | 11 | 0.761 |
| SUN | 10 | 0.738 |

We can describe the coming groups by a Binomial distribution with the parameters of Table 1.2. However this was only the first choice at the beginning of this paragraph, and looking back we find the alternative for the Binomial distributions. This is to take the given data of Table 1.1 (page 4) as a uniform distribution itself.

With such a uniform distribution the given occurrences all have equal chances, and all other occurrences have zero chance. Thus for the Fridays, the chance on welcoming 9 groups is *1/8*, the chance on 10 arriving groups is *0* and the chance on 11 interested groups is *1/8* again. This phenomenon appears six times in total, and doesn't seem to have any logical basis. Therefore we have chosen the first option. Although 8 measurements are few for deriving a distribution, we prefer this to an irrational fit by the data itself. In the next paragraph, the 7 Binomial distributions will form the basis for calculating the expected costs per year.

## 3.3 Stochastic analysis of improvements

We now want to recalculate the costs for different values of *n*, *k* and *f*. Instead of dealing with means as fixed values for the number of groups, we now have distributions for the number of groups. The deterministic calculations were done in the previous chapter. There, we found that changing the period length of the guides from 10 to 7 days was an important improvement. The shortening could lower costs by more than *f*30,000.=.

In this stochastic case the switching to a 7 days period reduces the personnel costs even more. For the present situation we find the costs to be *f*985,529,56. So now, we only will look for the period length of 7 days, and search for the optimal number of guides. Before calculating the expected costs, the following point requires attention. In the first chapter, a 4% sick leave was reported. From now on the chance on having sick guides unwilling to contaminate visitors is included. The increase in costs is about *4% x 200 days x f300 per day* = *f2400.*= per guide.

Starting with the number of guides in the present situation, we determine the adjusted expected costs for the different schedules:

$n = 7$ (length of a period)
$k = 5$ (number of consecutive days on duty for a regular guide)
$f = 8$ (number of regular forces)

The five best schedules appear to be:

| Schedule | | | | | | | Costs: |
|---|---|---|---|---|---|---|---|
| 3 | 1 | 4 | 0 | 0 | 0 | 0 | *f*900,072.50 |
| 3 | 2 | 3 | 0 | 0 | 0 | 0 | *f*900,192.50 |
| 3 | 1 | 3 | 0 | 1 | 0 | 0 | *f*902,731.94 |
| 3 | 2 | 3 | 0 | 1 | 0 | 0 | *f*903,061.62 |
| 3 | 1 | 3 | 0 | 0 | 1 | 0 | *f*903,324.12 |

The costs are definitely higher than in the deterministic model, even with regard to the expected increase by introducing sick leave. This is purely a result of working with stochastic input. The optimal schedule will depend on the average number of groups, but the variance occurs as a discrepancy in the tuning of the schedule to the groups. The discrepancy increases costs in both directions : having too many guides means that guides can remain idle, and too few guides requires part-timers.

Nevertheless, an increase of the number of guides leads to a decrease of costs, as in the deterministic approach. The optimal schedules of some increments can be omitted by only reporting the optimal costs:

| | | |
|---|---|---|
| $f = 9$, | Costs: | *f*823,501.19 |
| $f = 10$, | | *f*772,587.25 |
| $f = 11$, | | *f*744,270.19 |

Again, the optimal result appears for 12 guides:

$n = 7$ (length of a period)
$k = 5$ (number of consecutive days on duty for a regular guide)
$f = 12$ (number of regular forces)

The five best schedules appear to be:

| Schedule | | | | | | | Costs: |
|---|---|---|---|---|---|---|---|
| 5 | 1 | 5 | 0 | 0 | 1 | 0 | *f*735,233.87 |
| 4 | 2 | 4 | 0 | 1 | 1 | 0 | *f*735,679.25 |
| 4 | 2 | 5 | 0 | 0 | 1 | 0 | *f*735,951.06 |
| 4 | 1 | 5 | 0 | 1 | 1 | 0 | *f*736,002.00 |
| 5 | 1 | 5 | 0 | 1 | 0 | 0 | *f*736,045.00 |

For $f = 13$, costs increase again.

We find the optimal number of guides to be 12, as in the deterministic case. The optimal schedule is different, but in both cases most guides start working on Monday and Wednesday. An important aspect is the reduction of the costs. Although they are higher then in the simpler model, as expected, the costs fall down much faster when $f$ is increased to 12. As a result, the difference between the optima only is *f*45,000. The raise due to the introduction of sick leave is about *f*29,000, so the difference between deterministic and stochastic model is rather small.

# 4 Social aspects

## 4.1 Elaborating the schedules

When we look closer at the optimal schedule found in the previous section, we see that the costs may be optimised but some other things are not. The schedule found in the optimum for 12 guides is:

Schedule

5 1 5 0 0 1 0

The first number means 5 guides start working on Monday. With the period parameters (7,5) they work until Friday and are off in the weekend. Compare this with the 5 employees starting on Wednesday and you find that those guides are working in the weekend, while their colleagues are free. This injustice holds for the whole year, because the period length is a week. Every period length which multiples 7 will deal with this problem: guides starting on Monday have free weekends, starting on Sunday or Tuesday means half a weekend is available, but other starting days imply full working weekends.

Apart from the weekends, the schedule contains other social aspects. You can imagine a guide does not appreciate a working period of more than, say, 10 succeeding days. On the other hand, he is likely to reject irregular periods and to prefer an average ratio of working days and days off. For example, a period of 2 working days, 2 days off, 3 working days, 1 day off is unbalanced and split up too many times. It may seem silly to mention these social aspects for the length of a working period and the indicated ratio are constant in the observed schedule system. We need these aspects later on however, so we add them to the main point of weekends off.

As a result of the findings on the weekends, we have to elaborate the 7 days period schedules. If we can't solve the problem with injustice on the allocation, the choice for the 7 days has to be reconsidered.

## 4.2 Implementing the aspects

The main social problem with our optimal solution is the allocation of the free weekends. The problem itself is rather easy to solve. Let the starting days vary over the days of the period, instead of giving the guide a fixed starting day. Now the weekends off shift according to the starting days. However, this shifting affects the individual scheme for the year: beginning on Monday one next period and the next on Wednesday, means that you are off 4 days. Analogous, a guide can have to work for 10 days in a row. So, with the shiftings, the other social aspects mentioned in the previous paragraph arise.

Nevertheless, shifting the starting days is useful. Taking care for the balance of the individual working schemes is done by adding several restrictions. A guide works at most 6 consecutive days. After 6 days working the guide is at least 2 days off, this means after less consecutive working days 1 free day can be enough. Finally the leading aspect is implemented: the right allocation of weekends is guaranteed by requiring 1 free weekend in 4 weeks.

## 4.3 Shifting tables

The calculation of the expected personnel costs is extended with the developed social restrictions. As mentioned, this only was necessary for the 7 days working period. For several values of $f$, the number of guides, the effect of the restrictions is observed. However, the restriction has almost no influence on the optimal costs. For $f = 10$ and $f = 11$, several social shiftings for the optimal schedule were made. For $f = 12$ and $f = 13$, the second best optimal schedule offered social benefits towards the guides. While 12 guides is optimal and remains optimal for the second best schedule, our attention reaches only these calculations. The five best schedules appeared to be (as in chapter 3):

| Schedule | | | | | | | Costs: |
|---|---|---|---|---|---|---|---|
| 5 | 1 | 5 | 0 | 0 | 1 | 0 | *f*735,233.87 |
| 4 | 2 | 4 | 0 | 1 | 1 | 0 | *f*735,679.25 |
| 4 | 2 | 5 | 0 | 0 | 1 | 0 | *f*735,951.06 |
| 4 | 1 | 5 | 0 | 1 | 1 | 0 | *f*736,002.00 |
| 5 | 1 | 5 | 0 | 1 | 0 | 0 | *f*736,045.00 |

The adjustment to the second schedule gives an increase of the costs of about *f*400.=, this is less than 0.1%. So, there seems to be no threshold to prefer the second scheme. We do this,

and look at the corresponding shiftings. The social scheme list contains a list of feasible shiftings for a special working schedule. One scheme, the total shifting sequence, denotes the succeeding starting days. The number of starting days to shift over is depending on the number of guides. With 12 guides, the shiftings are notated in a series of 12 numbers.

| Costs | Schedule | | | | | | |
|---|---|---|---|---|---|---|---|
| *f*735,679.25 | 4 | 2 | 4 | 0 | 1 | 1 | 0 |

Social scheme list

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 3 | 2 | 1 | 3 | 3 | 2 | 1 | 5 | 6 | 3 | 1 |
| 1 | 3 | 2 | 1 | 3 | 3 | 2 | 1 | 6 | 5 | 3 | 1 |
| 1 | 3 | 2 | 1 | 5 | 3 | 2 | 1 | 6 | 3 | 3 | 1 |
| 1 | 3 | 2 | 1 | 5 | 6 | 2 | 1 | 3 | 3 | 3 | 1 |
| 1 | 3 | 2 | 1 | 6 | 3 | 2 | 1 | 5 | 3 | 3 | 1 |

Each number in the series denotes a starting day, and you will find each shifting contains 4 times day 1, 2 times day 2, 4 times day 3, etcetera. So, each guide follows the same shifting scheme, for example the first one, starting at 1 of the 12 places in the list. For initiation, each guide has to take a different place in the shifting, so that each place in the shifting scheme is filled up. The social requirements are fulfilled this way. Enough shifting sequences are available, because 32 different sequences correspond to the aspects. Therefore the problem of allocation of weekends with a 7 day period is solved, in a way the most important other social aspects are attended as well.

# 5 Recommendations

## 5.1 Improving the guide situation

The request for investigating the personnel costs was based on the assumption that the present situation can be improved. The fact it can be improved may be clear. The expected costs calculated with the stochastic model for the present situation amount to *f*985,530.=, whereas the costs after reorganization amount to *f*735,679.=. The first step in the reorganization is the employment of 4 new guides. Furthermore the working system of *7 days working - 3 days off* has to be changed in *5 days working - 2 days off*. These main changes lower personnel costs and can be applied within three important restrictions which hold the social aspects towards the guides.

To implement the new working system two schedules are necessary. The first schedule indicates for all days of the working period how many guides start to work their 5 days. The optimal schedule for starting days comes out of a few thousand schedules, and therefore the difference in costs between optimal and 100$^{th}$ best is less than 2%. So, a lot of alternatives almost equal the optimal schedule. We choose the second-best solution, because this schedule satisfied the social requirements. The schedule is stated as *4 2 4 0 1 1 0* in the previous chapter, which means:

- 4 guides start to work on Monday,
- 2 on Tuesday,
- 4 on Wednesday,
- 1 on Friday,
- 1 on Saturday.

The second scheme guarantees the guides have the same amount of free weekends. Each guides follows a sequence of 12 starting days, each period he starts on a different day and after 12 periods he follows the sequence again. Although 32 schemes fit, we present only one for simplicity. The first sequence of the *Social scheme list* in the previous section produces the series:

Monday - Wednesday - Tuesday - Monday - Wednesday - Wednesday - Tuesday - Monday - Friday - Saturday - Wednesday - Monday.

Each of the 12 days is occupied by a guide. The guide follows the series until the end and then runs the sequence from the first Monday on continuously.

## 5.2 Costs and social aspects

The solution as presented in the previous paragraph has some limitations in both uniqueness and optimisation procedure. Despite the optimal solution fit to the social restrictions is unique, many alternative options produce almost the same costs. The differences are only fractions of a percent. The alternatives do in fact not differ in number of guides and working period, but only diverge in the working schedule and social scheme for shifting. The limitation of the uniqueness is however a benefit of the solution. We can choose out of a range of schedules and shiftings with only extremely small influence on costs.

The other limitation has to do with the relevance of the improvement. The personnel employment is optimised to the personnel costs. These costs include only the salaries of the guides, part-timers and scientific staff. The improvements have indeed effect on non-financial aspects such as satisfaction of the guides. The reduction of the working period and the introduction of shiftings affect social aspects, this can be in positive as well in negative way. We prefer to include these impacts in the optimisation by expressing them in terms of costs and benefits. This is however an extremely difficult task, and therefore disregarded in this model. The least you can do is to bare in mind these impacts when comparing the different improvements. The last remark concerning the relevance is about the sensibility on the data. The improvements for personnel employment depend on the numbers of arriving groups. The amount of supplied data is small, which is a problem for estimating a distribution. When looking at the difference between deterministic and stochastic models you can see the influence of the input data is significant. We therefore recommend that the model should be reused when more figures about groups are available.

# Appendix A Statistical testing

The Chi Square Test for Goodness of Fit is perhaps the best known statistical test. It tests the hypothesis a sample of data is coming from a supposed distribution. The test parameter $Y$ in formula (3) has, as may be deduced from its name, a chi-square distribution.

$$Y = \sum_{j=1}^{k} \frac{(N_j - N \times p_j)^2}{N \times p_j} \quad , \qquad \sum_{j=1}^{k} N_j = N \tag{3}$$

The total range of outcomes of the tested distribution is divided in $k$ classes. $N_j$ is the umber of outcomes in class $j$, and $N$ is the total number of measurements. The Binomial distribution with parameters $(n,k)$ has outcomes $0, 1, .., n$. Therefore, we make $n+1$ classes, one for each integer outcome. Table 1.3 contains two parameter sets for every day, the calculated test values and a column for calculated squares.

**Table 1.3** Chi Square Test for the Binomial distribution

| Day | n | p | Squares | Chi-Squares |
|---|---|---|---|---|
| MON | 8 | 0.828 | 2.49 | 3.58 |
| | 9 | 0.736 | 2.40 | 2.51 |
| TUE | 10 | 0.850 | 1.45 | 1.97 |
| | 11 | 0.773 | 0.92 | 1.57 |
| WED | 15 | 0.900 | 1.43 | 1.83 |
| | 16 | 0.844 | 0.88 | 1.38 |
| THU | 11 | 0.898 | 2.49 | 6.61 |
| | 12 | 0.823 | 1.47 | 2.91 |
| FRI | 16 | 0.820 | 4.66 | 13.2 |
| | 17 | 0.772 | 4.8 | 17.4 |
| SAT | 10 | 0.837 | 1.64 | 9.46 |
| | 11 | 0.761 | 1.59 | 5.58 |
| SUN | 10 | 0.738 | 2.86 | 6.50 |
| | 11 | 0.670 | 3.64 | 13.1 |

These squares are the squared differences of measurement and predicted outcome, whereas the chi-squares are squared differences with weight on the predicted outcome. The squares are added to compare with the chi-squares. The function of the last values is to determine the best out of two parameter sets for describing the number of groups. The best set is the one with lowest test value. In general, the test value is used to reject or not reject a parameter (set). The rejection takes place when the $Y$ exceeds a critical value. While $Pr(\chi^2_7<14) = 0.95$, the critical value is $14$ with a significance of $5\%$ and $7$ degrees of freedom. (Degrees of freedom is number of measurements minus one.)

Most of the test values in Table 1.3 fall below this boundary. This is however not what we are looking at. The purpose of the test is to choose the sets for Table 1.2 on page 15. For five days the highest $n$ appeared to be best, harmonizing with the lowest $p$. In all decisions the lowest chi-squares matched with the lowest squares. The meaning of the critical value is not used. In fact the test needs far more data to be reliable.

# DESIGN OF AN IRRIGATION SYSTEM

# Contents

## Abstract

An irrigation system has to be designed for a rectangular area with a water source in one of the corners. Of course, there is a trade-off between investment and maintenance costs for the irrigation system and savings in transportation costs using this system. The only restriction for such an irrigation system is that it has to consist of a number of parallel canals, that are connected to the source by a pipeline. This still leaves a lot of freedom for the design. Several types of irrigation system designs are considered, where for each type of design a number of parameters has to be specified. The parameter values yielding lowest costs are found numerically. Different designs are compared to each other with respect to the total costs over a period of 10 years, which seems to be a reasonable period to consider. The cheapest irrigation system that has been found gives savings of approximately 99% over 10 years.

# 1 Introduction

This report concerns a project studied for the modelling colloquium for students in the post-graduate programme "Mathematics for Industry". The problem is called "Irrigation System" and was worked on from half of March to the end of April 1993.

The problem concerns a rectangular area that has to be supplied with water. This water comes from a source that is situated in one of the corners of the area. The costs of transportation of water are relatively high. Therefore it is assumed that it will be useful to construct an irrigation system of parallel canals. The aim of this project is to find a good, i.e. a cost effective, system of canals. Different systems can be compared since all costs are given.

The irrigation system has to consist of parallel canals that are connected to the source by a pipeline. Several designs are possible. In each of these designs, the values of some parameters, such as the distance between canals, are not yet specified. Optimal values for these parameters are found numerically, and the different designs are compared.

In the next section, the problem description will be given, together with the introduction of assumptions and variables. Then, in section 3, the costs for the current situation (without an irrigation system) are computed. This allows us to determine the savings if an irrigation system is applied. In section 4, several designs for irrigation systems are considered and optimal values for the parameters are computed. Finally, a best design is selected. In section 5, conclusions are drawn.

# 2 Problem description

## 2.1 Data and assumptions

For agricultural purposes, a rectangular area has to be supplied with water. The farmers in the area get the water they need from a source that is in one of the corners. It is assumed that the farmers take the shortest route, i.e. a straight line, to the source. Since no information is available about the distribution of the need for water, it is assumed that there is a uniform demand.

This is quite an expensive way of transporting the water. Therefore, it is suggested to construct a system of parallel canals in the area, so that the farmers can get water from these canals. The water will be pumped from the source to the canals through a pipeline. It is assumed that canals, pipelines and pumps can be put everywhere in the area and have negligible width. If there is an irrigation system, the farmers get water from the nearest canal.

Several costs are involved. First there are the costs for transportation of water from the nearest canal (or in the current situation, from the source) to the place where the water is needed. Then, if an irrigation system is constructed, there are costs for digging the canals and yearly costs for their maintenance. Furthermore, there are costs for constructing the pipeline and yearly costs for its maintenance. The water has to be pumped through the pipeline, and it is assumed that a pump can only pump water over a certain range. If the length of the pipeline is larger than this range, more pumps will have to be installed. The pumps also have investment costs and yearly maintenance and energy costs. It is assumed that these costs depend linearly on the amount of water that has to be pumped. The first pump is a large pump, since it has to pump all the water that is needed. Every next pump can be smaller and therefore have lower costs. All costs and other values are given in the next table.

| | | |
|---|---|---|
| $l$ | length of the area | 12 000 m |
| $b$ | width of the area | 10 000 m |
| $w$ | water needed | 3 000 liter per $m^2$ per year |
| $l_p$ | reach of one pump | 5 000 m pipeline |
| $c_t$ | costs for transport of water | Dfl $10^{-5}$ per liter per m |
| $c_{\text{invcanal}}$ | costs for making the canals | Dfl 300 per m canal |
| $c_{\text{yearcanal}}$ | maintenance of the canals | Dfl 30 per m canal per year |
| $c_{\text{invpipe}}$ | costs for making the pipeline | Dfl 2 000 per m pipeline |
| $c_{\text{yearpipe}}$ | yearly costs for the pipeline | Dfl 100 per m pipeline per year |
| $c_{\text{invpump}}$ | costs for making a pump | Dfl 250 000 per pump |
| $c_{\text{yearpump}}$ | yearly costs for a pump | Dfl 10 000 per pump per year |
| $p$ | interest factor | 1.09 (9% interest) |

## 2.2 "Upper", "lower", "left" and "right"

In this report, the rectangular area is described as being in landscape orientation, with the source in the lower left corner (see Figure 1). For ease

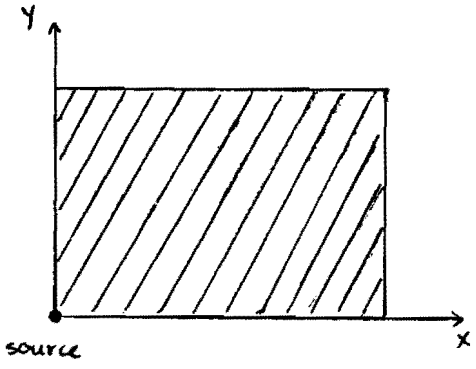



Figure 1: The area.

of notation, we will indicate directions and borders with "upper", "lower", "left" and "right". Of course, the results do not depend on the orientation. For "the left border" one should in fact read "the short border at the side of the source", etc.

## 2.3 The cost functions

The purpose of this project is to design an irrigation system with minimal costs. Since there are costs for the investment and yearly costs we will

have to consider a certain period, for example 10 years (the period in which the canals are written off for depreciation). How this can be done will be explained in Section 2.4.

First we will introduce some variables. A list of all used variables is given in Appendix A. The total length of the canals will be called $x_c$ and the length of the pipeline $x_p$. The average distance to get water (in the current situation the distance to the source, in the situation with the irrigation system the distance to the nearest canal) will be called $\bar{r}$. This gives transportation costs $c_{\text{transport}} = lbwc_t\bar{r}$ per year[1]. The number of pumps is called $N_{\text{pumps}}$ and equals $\left\lfloor \frac{x_p}{l_p} \right\rfloor$, where $\lfloor \cdot \rfloor$ denotes rounding off downwards. The costs $c_{\text{invpump}}$ and $c_{\text{yearpump}}$ are costs for the first pump, i.e. the pump that has to pump all the water. All next pumps have costs proportional to the length of the pipeline they have to serve, so the total investment costs for the pumps are (see Figure 2)

$$\sum_{i=0}^{N_{\text{pumps}}-1} c_{\text{invpump}} \frac{x_p - il_p}{x_p}$$

and the yearly pumping costs are computed similarly.

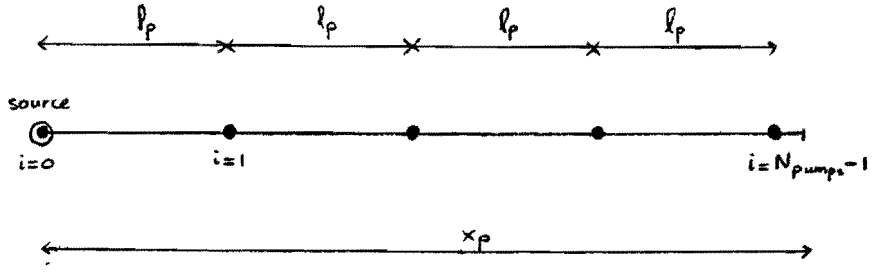



Figure 2: The pump distances.

Summarizing, the costs that are involved are

Investment costs:

- Canals : $c_{\text{invcanal}}x_c$
- Pipeline : $c_{\text{invpipe}}x_p$

[1]In fact, most of our computations do not use this formula for the entire area, but it is used for parts or we integrate over an area.

- Pumps : $\sum_{i=0}^{N_{\text{pumps}}-1} c_{\text{invpump}} \frac{x_p - il_p}{x_p}$

Yearly costs:

- Transport of water : $c_{\text{transport}} = lbwc_t\bar{r}$
- Maintenance of the canals : $c_{\text{yearcanal}} x_c$
- Maintenance of the pipelines : $c_{\text{yearpipe}} x_p$
- Pump maintenance and pumping costs : $\sum_{i=0}^{N_{\text{pumps}}-1} c_{\text{yearpump}} \frac{x_p - il_p}{x_p}$
- Interest (see section 2.4)

## 2.4 The interest factor

Before we can compute costs for irrigation systems, we have to say something about the time period to be considered and the interest involved. We will consider a period of $N$ years, where $N = 10$ since that is the period in which the canals are written off for depreciation. The total yearly and investment costs are called $c_{\text{year}}$ and $c_{\text{inv}}$, respectively, and the interest factor is $p$ (in our case, this factor is 1.09, since the interest is 9%). Now we can compute the costs after $N$ years, including interest. The investment was done $N$ years ago, so inlcuding the loss of interest, the costs after $N$ years are $p^N c_{\text{inv}}$. All years after that, there were yearly costs $c_{\text{year}}$, so including interest, the costs of $k$ years ago are $p^k c_{\text{year}}$. The total costs are:

$$\begin{aligned} c_{\text{total}} &= p^N c_{\text{inv}} + c_{\text{year}} + pc_{\text{year}} + p^2 c_{\text{year}} + \cdots + p^{N-1} c_{\text{year}} \\ &= p^N c_{\text{inv}} + \frac{1-p^N}{1-p} c_{\text{year}}. \end{aligned}$$

If we want to compare costs for different irrigation systems, we divide the total cost function by $\frac{1-p^N}{1-p}$ and consider

$$c = c_{\text{year}} + f c_{\text{inv}},$$

where $f$ is a factor equal to

$$f = \frac{1-p}{1-p^N} p^N.$$

For $N = 10$ the factor $f$ is approximately equal to 0.156. The costs $c$ are a measure for the costs per year, viewing over a period of $N$ years, and they can also be compared with the yearly costs in the current situation, that will be computed in section 3.

# 3 The current situation

In this section the yearly costs for the current situation will be computed. To do this, we will have to compute the average distance to the source. This distance is equal to

$$\bar{r} = \frac{1}{lb}\int_{x=0}^{l}\int_{y=0}^{b}\sqrt{x^2+y^2}\mathrm{d}x\mathrm{d}y.$$

To compute this distance, the area will be divided into two parts, A and B (see Figure 3), and we will use polar coordinates.

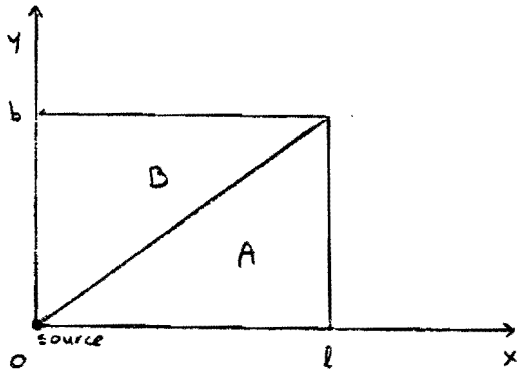



Figure 3: Division of the area.

For part A, the integral is transformed into

$$\int_{\phi=0}^{\arctan\frac{b}{l}}\int_{r=0}^{\frac{l}{\cos\phi}} r^2\mathrm{d}r\mathrm{d}\phi,$$

which equals

$$\frac{1}{3}\int_{\phi=0}^{\arctan\frac{b}{l}}\frac{l^3}{\cos^3\phi}\mathrm{d}\phi = \frac{1}{3}l^3\int_{\phi=0}^{\arctan\frac{b}{l}}\frac{\cos\phi}{(1-\sin^2\phi)^2}\mathrm{d}\phi = \frac{1}{3}l^3\int_{y=0}^{\frac{b}{\sqrt{b^2+l^2}}}\frac{1}{(1-y^2)^2}\mathrm{d}y,$$

and for part B, the integral becomes

$$\int_{\phi=\arctan\frac{b}{l}}^{\frac{\pi}{2}} \int_{r=0}^{\frac{b}{\sin\phi}} r^2 \mathrm{d}r\mathrm{d}\phi = \frac{1}{3}b^3 \int_{\phi=\arctan\frac{b}{l}}^{\frac{\pi}{2}} \frac{1}{\sin^3\phi}\mathrm{d}\phi = \frac{1}{3}b^3 \int_{x=0}^{\frac{l}{\sqrt{b^2+l^2}}} \frac{1}{(1-x^2)^2}\mathrm{d}x.$$

Consequently, the average distance is

$$\bar{r} = \frac{1}{3lb}\left[b^3 \int_{x=0}^{\frac{l}{\sqrt{b^2+l^2}}} \frac{1}{(1-x^2)^2}\mathrm{d}x + l^3 \int_{y=0}^{\frac{b}{\sqrt{b^2+l^2}}} \frac{1}{(1-y^2)^2}\mathrm{d}y\right],$$

which gives

$$\bar{r} = \frac{1}{3}\sqrt{b^2+l^2} + \frac{b^2}{12l}\ln\left(\frac{\sqrt{b^2+l^2}+l}{\sqrt{b^2+l^2}-l}\right) + \frac{l^2}{12b}\ln\left(\frac{\sqrt{b^2+l^2}+b}{\sqrt{b^2+l^2}-b}\right).$$

For the given values, this equals approximately 8.4 km. Since

$$c_{\text{transport}} = lbwc_t\bar{r} = \text{Dfl } 3.6 \cdot 10^6\bar{r},$$

the costs per year are Dfl $3.04 \cdot 10^{10}$.

# 4 Irrigation system designs and cost optimization

In general, the irrigation system may be very complex. The only restriction is that the canals are parallel and connected to the source by a pipeline. It is not possible to make computations for such a general system, so we will consider a smaller class of irrigation systems, that we expect to give good results. First we will consider canals that are parallel to one of the borders of the area and run from one border to the other. Our next step will be to design a system of canals that are still parallel to one of the borders, but do not extend from one border to another. Finally, we will consider a system with canals that are not parallel to one of the borders.

A different approach might be to dig a lot of short parallel canals in line, but since the canals have to be connected by pipelines, which are quite expensive, this will not be optimal, so we will not consider such systems.

## 4.1 Canals from one border to the other

In this section we consider a simple system: canals that are parallel to one of the borders and extend from one border to the other. This system is illustrated in Figure 4.

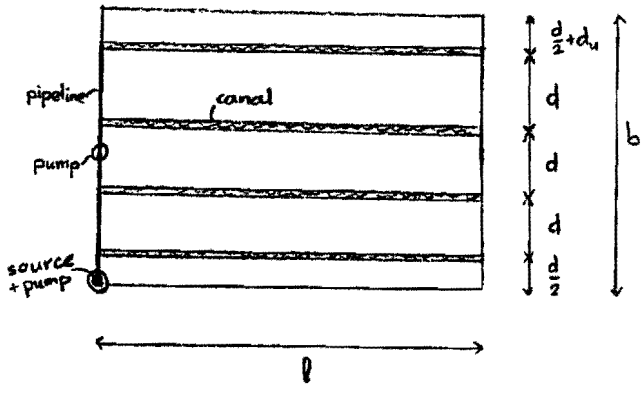



Figure 4: Canals from one border to the other.

Other choices include:

- The canals are parallel to the longest border instead of perpendicular to it.

- The distances between canals are equal ($d$).
- The distance between the lower border and the first canal is $\frac{d}{2}$.
- The distance between the upper border and the last canal is larger than $\frac{d}{2}$, namely $\frac{d}{2} + d_u$.

First we will explain the choice of the direction of the canals. Assume that a certain distance between canals, or a certain average transportation distance, is desired. Then, for the total length of *canals*, it makes no difference whether the canals are parallel to the longest or to the shortest border (in one case the canals are longer, in the other case there are more canals). However, the *pipeline* will be shorter if the canals are parallel to the longest border, so that will be our choice.

The choices for the *distances* were made for the following reasons. If there were only canals, and no pipelines, the cheapest way of putting $n$ canals in the area would be to spread them out equally, i.e. to put them at an equal distance $d$ from each other and the first and the last canal at distances $\frac{d}{2}$ from the borders. It is easy to check that in this manner, transportation costs are lowest. However, there has to be a pipeline from the source, connecting all canals. This pipeline along the left border will extend from the source to the last canal (and not to the border, see Figure 4). Pipelines are quite expensive, so we make the distance from the last canal to the border larger than $\frac{d}{2}$.

There are two parameters that have to be chosen, namely the number of canals $n$ and the extra distance from the last canal to the border, which will be $d_u$. The number of canals is related to the distance between the canals, since

$$d = \frac{b - d_u}{n}.$$

The total length of canals equals $x_c = nl$ and the length of the pipeline is $x_p = (n - \frac{1}{2})d$. Therefore, the number of pumps is

$$N_{\text{pumps}} = \left\lfloor \frac{x_p}{l_p} \right\rfloor = \left\lfloor \frac{(n - \frac{1}{2})d}{l_p} \right\rfloor.$$

Now we can compute the costs as a function of the two variables $n$ and $d_u$.

The investment costs are

$$\begin{aligned} c_{\text{inv}} &= c_{\text{invcanal}} x_c + c_{\text{invpipe}} x_p + \sum_{i=0}^{N_{\text{pumps}}-1} c_{\text{invpump}} \frac{x_p - i l_p}{x_p} \\ &= c_{\text{invcanal}} n l + c_{\text{invpipe}} (n - \frac{1}{2}) d + \sum_{i=0}^{N_{\text{pumps}}-1} c_{\text{invpump}} \frac{(n - \frac{1}{2})d - i l_p}{(n - \frac{1}{2})d} \end{aligned}$$

and the yearly costs are

$$\begin{aligned} c_{\text{year}} &= c_{\text{yearcanal}} x_c + c_{\text{yearpipe}} x_p + \sum_{i=0}^{N_{\text{pumps}}-1} c_{\text{yearpump}} \frac{x_p - i l_p}{x_p} + c_{\text{transport}} \\ &= c_{\text{yearcanal}} n l + c_{\text{yearpipe}} (n - \frac{1}{2}) d + \sum_{i=0}^{N_{\text{pumps}}-1} c_{\text{yearpump}} \frac{(n - \frac{1}{2})d - i l_p}{(n - \frac{1}{2})d} \\ &\quad + w c_t ((n - \frac{1}{2}) d l \frac{d}{4} + (\frac{d}{2} + d_u) l \frac{\frac{d}{2} + d_u}{2}), \end{aligned}$$

where $N_{\text{pumps}} = \left\lfloor \frac{(n-\frac{1}{2})d}{l_p} \right\rfloor$ and $d = \frac{b - d_u}{n}$. The last term in the formula can be explained as follows: $(n - \frac{1}{2})dl$ is the area between the lower border and the last canal. In this area the average distance to the nearest canal is $\frac{d}{4}$. The area between the last canal and the "upper" border is $(\frac{d}{2} + d_u)l$, and in this area the average distance is $\frac{\frac{d}{2} + d_u}{2}$.

All variables are given except $n$ and $d_u$, so we can minimize the costs expression as a function of $n$ and $d_u$. This was done numerically with a simple local search method.

The optimum that has been found is

$$\begin{aligned} n &= 99 \\ d_u &= 1.2 \text{ m} \\ d &= 101.0 \text{ m} \\ c &= \text{Dfl } 186.2 \cdot 10^6 \end{aligned}$$

Compared with the costs in the current situation (Dfl $3.04 \cdot 10^{10}$), there are very large savings if this irrigation system is constructed.

## 4.2 Canals that do not extend to the right border

Now that an optimum is found for a system of canals extending from one border to the other, we will generalize our model to canals that do not extend to the right border. This will probably be cheaper, since the costs for the canals decrease linearly while the transportation distances increase less.

In Figure 5 this system is depicted. The distance from the right end of the canals to the right border will be called $d_r$.

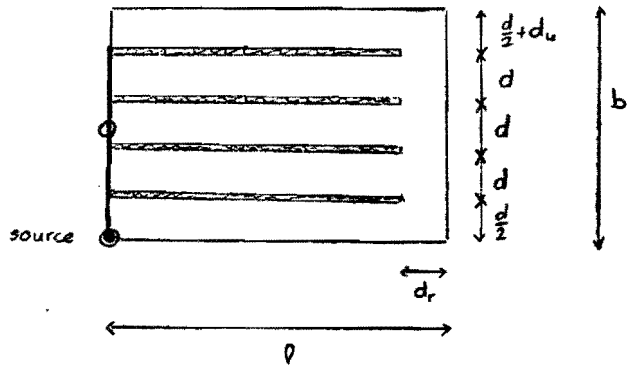



Figure 5: Canals that do not extend to the right border.

Now the costs have to be optimized with respect to three variables: $n$, $d_u$ and $d_r$. The cost function is more complex than in the previous case, since the distances from points in the area right to the nearest canal are more difficult to compute. In Appendix B the cost function is derived. The optimal values are

$$
\begin{aligned}
n &= 99 \\
d_u &= 1.2 \text{ m} \\
d_r &= 42.0 \text{ m} \\
d &= 101.0 \text{ m} \\
c &= \text{Dfl } 186.1 \cdot 10^6
\end{aligned}
$$

We see that the costs are slightly lower, so this system is a little better than the system of the previous section. However, the differences are small, and the found parameter values are nearly the same.

## 4.3 Canals that do not extend from one border to the other

The same argument as in the previous section can be used to make a system of canals that do not extend to the right, nor to the left border. In Figure 6 this system is illustrated. In this case the pipeline looks different (it has a bend), and there is no reason to make the distance from the lower border to the first canal equal to $\frac{d}{2}$, since the pipeline influences the costs. Therefore we have to define the distances $d_u$, $d_d$, $d_r$ and $d_l$ in addition to $d$.

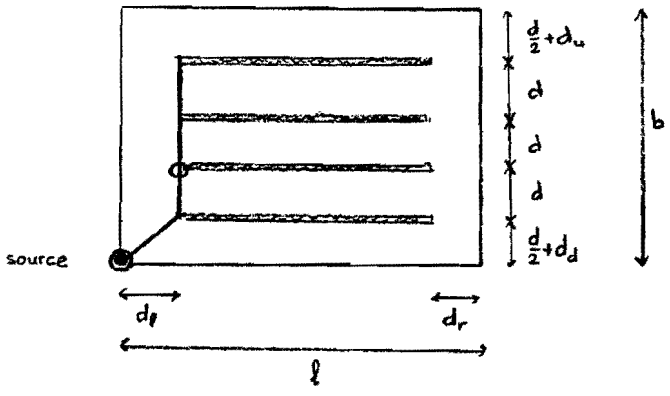



Figure 6: Canals that do not extend from one border to the other.

Again the cost function is given in Appendix B. The optimum turns out to be

$$
\begin{aligned}
n &= 99 \\
d_u &= 1.2 \text{ m} \\
d_d &= 0.3 \text{ m} \\
d_r &= 42.0 \text{ m} \\
d_l &= 40.9 \text{ m} \\
d &= 101.0 \text{ m} \\
c &= \text{Dfl } 185.9 \cdot 10^6
\end{aligned}
$$

We see that the costs are lower again, so this irrigation system is better than the other two, although the differences are not very large. This system has the property that in some areas one has to cross the pipeline to get to the nearest canal. If this turns out to be a disadvantage, one might build one of the other systems, since the costs are nearly equal.

Looking at the found optimal values, one can see that the distances from the first and the last canal to the borders do not differ very much from $\frac{d}{2}$. The number of canals is 99 in all cases, and the distances from the canals to the left and right borders are almost equal. The fact that the distance to the left border is smaller than the distance to the right border can be explained by the fact that there has to be a pipeline from the corner to the first canal, which makes it expensive to have the canals far away from the left border.

## 4.4 Canals that are not parallel to a border

In the previous sections we considered only canals that were parallel to one of the borders. To see if it will be cheaper if they are not, we will consider a design with canals that are at a certain angle to the longest border. This design is illustrated in Figure 7. The design of the pipeline is

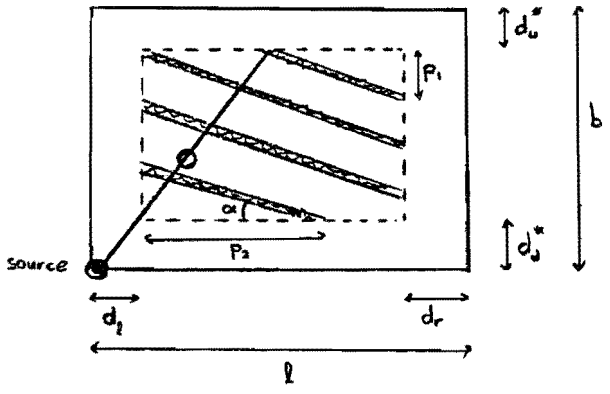



Figure 7: Canals that are not parallel to one of the borders.

more complicated, since it has to be as short as possible with the requirement that it has to connect all canals to the source. If possible, we make it perpendicular to the canals, otherwise it will run form the source to one of the endpoints of the last canal or along one of the endpoints of the first canal. It may even be necessary to make a pipeline with a bend. We will not treat this in detail. There are 7 parameters to determine, namely the

distances from the borders[2] $d_l$, $d_r$, $d_d^*$ and $d_u^*$, the number of canals $n$, the angle between the canals and the lower border $\alpha$ and two parameters $p_1$ and $p_2$ that determine where the first and the last canal are. The optimal values are

$$\begin{aligned} n &= 99 \\ d_u^* &= 51.4 \text{ m} \\ d_d^* &= 50.8 \text{ m} \\ d_r &= 41.9 \text{ m} \\ d_l &= 40.9 \text{ m} \\ \alpha &= 0^\circ \\ p_1 &= 0.2 \text{ m} \\ p_2 &= 11916.2 \text{ m} \\ c &= \text{Dfl } 185.9 \cdot 10^6 \end{aligned}$$

We see that the solution degenerates into the solution of the previous section: canals parallel to one of the borders (small differences may be explained by numerical errors). Therefore we may conclude that it is no use to make canals that are not parallel to one of the borders.

[2] We used $d_d^*$ and $d_u^*$ instead of $\frac{d}{2} + d_d$ and $\frac{d}{2} + d_u$, since there is no reason anymore to make these distances approximately $\frac{d}{2}$.

# 5 Conclusions

It is very cost effective to construct an irrigation system for the given area. The costs will decrease approximately 99% (!) for a period of 10 years. Of the designs that were examined, the best one is a system of canals parallel to one of the borders, but not extending from one border to the other. This system is described in section 4.3 and optimal values for the parameters are computed numerically. The differences between this design and the other designs that were examined are not very large, compared with the savings. Therefore, if for some reason one prefers e.g. a system with canals from one border to the other, one may choose that as well. It might be possible that there are even cheaper designs, since not all possible configurations of parallel canals were tested, but looking at the small differences between the tested designs, we do not expect them to be much cheaper.

During the numerical computations, it turned out that the cost funtion is not very sensitive to changes in the parameters. If a system with canals at a distance of e.g. 95 m is constructed, or if the distances to the borders are smaller or larger than indicated in this report, this will not give much higher costs.

# A List of variables

| | | |
|---|---|---|
| $\alpha$ | | angle between the lower border and the canals |
| $b$ | m | width of the area |
| $c$ | Dfl year$^{-1}$ | yearly costs including interest over $N$ years |
| $c_{\text{inv}}$ | Dfl | total investment costs |
| $c_{\text{invcanal}}$ | Dfl m$^{-1}$ | investment costs for canals |
| $c_{\text{invpipe}}$ | Dfl m$^{-1}$ | investment costs for a pipeline |
| $c_{\text{invpump}}$ | Dfl | investment costs for a large pump |
| $c_t$ | Dfl l$^{-1}$ m$^{-1}$ | costs for transport of water |
| $c_{\text{total}}$ | Dfl | total costs over $N$ years |
| $c_{\text{transport}}$ | Dfl year$^{-1}$ | total yearly transportation costs |
| $c_{\text{year}}$ | Dfl year$^{-1}$ | total yearly costs |
| $c_{\text{yearcanal}}$ | Dfl m$^{-1}$ year$^{-1}$ | yearly costs for canals |
| $c_{\text{yearpipe}}$ | Dfl m$^{-1}$ year$^{-1}$ | yearly costs for the pipeline |
| $c_{\text{yearpump}}$ | Dfl year$^{-1}$ | yearly costs for a large pump |
| $d$ | m | distance between canals |
| $d_d$ | m | extra distance between lower border and first canal |
| $d_d^*$ | m | distance between lower border and first canal |
| $d_l$ | m | distance between left border and the canals |
| $d_r$ | m | distance between right border and the canals |
| $d_u$ | m | extra distance between upper border and last canal |
| $d_u^*$ | m | distance between upper border and last canal |
| $f$ | year$^{-1}$ | balancing factor for investment and yearly costs |
| $l$ | m | length of the area |
| $l_p$ | m | reach of one pump |
| $n$ | | number of canals |
| $N$ | year | considered time period |
| $N_{\text{pumps}}$ | | number of pumps |
| $p$ | | interest factor |
| $p_1$ | m | distance parameter in design with canals under an angle |
| $p_2$ | m | distance parameter in design with canals under an angle |
| $\bar{r}$ | m | average distance to the source or nearsest canal |
| $w$ | l m$^{-2}$ year$^{-1}$ | amount of water needed |
| $x_c$ | m | total length of canals |
| $x_p$ | m | length of the pipeline |

# B Cost functions

The costs that will be compared are

$$c = c_{\text{year}} + f c_{\text{inv}}.$$

For all designs, the investment costs are

$$c_{\text{inv}} = c_{\text{invcanal}} x_c + c_{\text{invpipe}} x_p + \sum_{i=0}^{N_{\text{pumps}}-1} c_{\text{invpump}} \frac{x_p - i l_p}{x_p}$$

and the yearly costs are

$$c_{\text{year}} = c_{\text{yearcanal}} x_c + c_{\text{yearpipe}} x_p + \sum_{i=0}^{N_{\text{pumps}}-1} c_{\text{yearpump}} \frac{x_p - i l_p}{x_p} + c_{\text{transport}},$$

where $N_{\text{pumps}} = \left\lfloor \frac{x_p}{l_p} \right\rfloor$. Different designs use different formulas for $x_c$ and $x_p$ and, most difficult to compute, $c_{\text{transport}}$. For each of the designs the formulas will now be given.

## B.1 Canals from one border to the other

The simple design of canals parallel to one of the borders, and extending from one border to the other is described in section 4.1 and illustrated in Figure 4. The formulas are already derived in section 4.1 and are

$$\begin{aligned} x_c &= nl \\ x_p &= (n - \frac{1}{2})d \\ c_{\text{transport}} &= wc_t \left\{ (n - \frac{1}{2}) dl \frac{d}{4} + (\frac{d}{2} + d_u) l \frac{\frac{d}{2} + d_u}{2} \right\}. \end{aligned}$$

## B.2 Canals that do not extend to the right border

The second design, canals that do not extend to the right border, is described in section 4.2 and illustrated in Figure 5. The formulas are:

$$
\begin{aligned}
x_c &= n(l - d_r) \\
x_p &= (n - \frac{1}{2})d \\
c_{\text{transport}} &= wc_t \left\{ (n - \frac{1}{2})d(l - d_r)\frac{d}{4} + (\frac{d}{2} + d_u)(l - d_r)\frac{\frac{d}{2} + d_u}{2} \right\} \\
&\quad + (2n + 1)RC(\frac{d}{2}, d_r) + RC(\frac{d}{2} + d_u, d_r).
\end{aligned}
$$

In the formula for the transportation costs, a function $RC$ occurs. This function $RC(h, l)$ determines the costs for a rectangular area of size $h \times l$ with the "source" (the end of the nearest canal) in one of the corners. This function is closely related to the formula that was derived in section 3:

$$
\begin{aligned}
RC(h, l) &= wc_t \left\{ \frac{1}{3}hl\sqrt{h^2 + l^2} + \frac{h^3}{12} \ln \left( \frac{\sqrt{h^2 + l^2} + l}{\sqrt{h^2 + l^2} - l} \right) \right. \\
&\quad \left. + \frac{l^3}{12} \ln \left( \frac{\sqrt{h^2 + l^2} + h}{\sqrt{h^2 + l^2} - h} \right) \right\}.
\end{aligned}
$$

## B.3 Canals that do not extend from one border to the other

This design is described in section 4.3 and illustrated in Figure 6. The formulas are

$$
\begin{aligned}
x_c &= n(l - d_r - d_l) \\
x_p &= (n - 1)d + \sqrt{d_l^2 + (\frac{d}{2} + d_d)^2} \\
c_{\text{transport}} &= wc_t \left\{ (n - 1)d(l - d_r - d_l)\frac{d}{4} + (\frac{d}{2} + d_u)(l - d_r - d_l)\frac{\frac{d}{2} + d_u}{2} \right.
\end{aligned}
$$

$$+(\frac{d}{2}+d_d)(l-d_r-d_l)\frac{\frac{d}{2}+d_d}{2}\Big\}+(2n-2)RC(\frac{d}{2},d_r)$$

$$+(2n-2)RC(\frac{d}{2},d_l)+RC(\frac{d}{2}+d_u,d_r)+RC(\frac{d}{2}+d_u,d_l)$$

$$+RC(\frac{d}{2}+d_d,d_r)+RC(\frac{d}{2}+d_d,d_l),$$

where $RC$ is the cost function for a rectangular area with the "source" in one of the corners (see section B.2).

## B.4 Canals that are not parallel to a border

If the canals are not parallel to one of the borders, the formulas become very complicated. First the pipeline has to be designed to make it possible to determine its length. We already mentioned the design of the pipeline in section 4.4. Then the transportation costs have to be computed. We will not give the exact formulas, but we will confine ourselves to saying that the area can be divided into regions with different structures. In Figure 8 these regions are given. A bold line indicates a canal along the border of a region, a point indicates the end of a canal, which can be considered as a "source". In a region with more than one place where one can get water has to be separated into a part where one gets water from one canal/"source" and a part where one gets water from the other canal/"source". All lengths and angles are also given in the figure. The total area can be divided into the following regions:

- A square with a "source" in one of the corners. The formula for the costs in such a region:

$$RC(h,l) = wc_t\left\{\frac{1}{3}hl\sqrt{h^2+l^2}+\frac{h^3}{12}\ln\left(\frac{\sqrt{h^2+l^2}+l}{\sqrt{h^2+l^2}-l}\right)\right.$$
$$\left.+\frac{l^3}{12}\ln\left(\frac{\sqrt{h^2+l^2}+h}{\sqrt{h^2+l^2}-h}\right)\right\}$$

(already mentioned in section B.2).

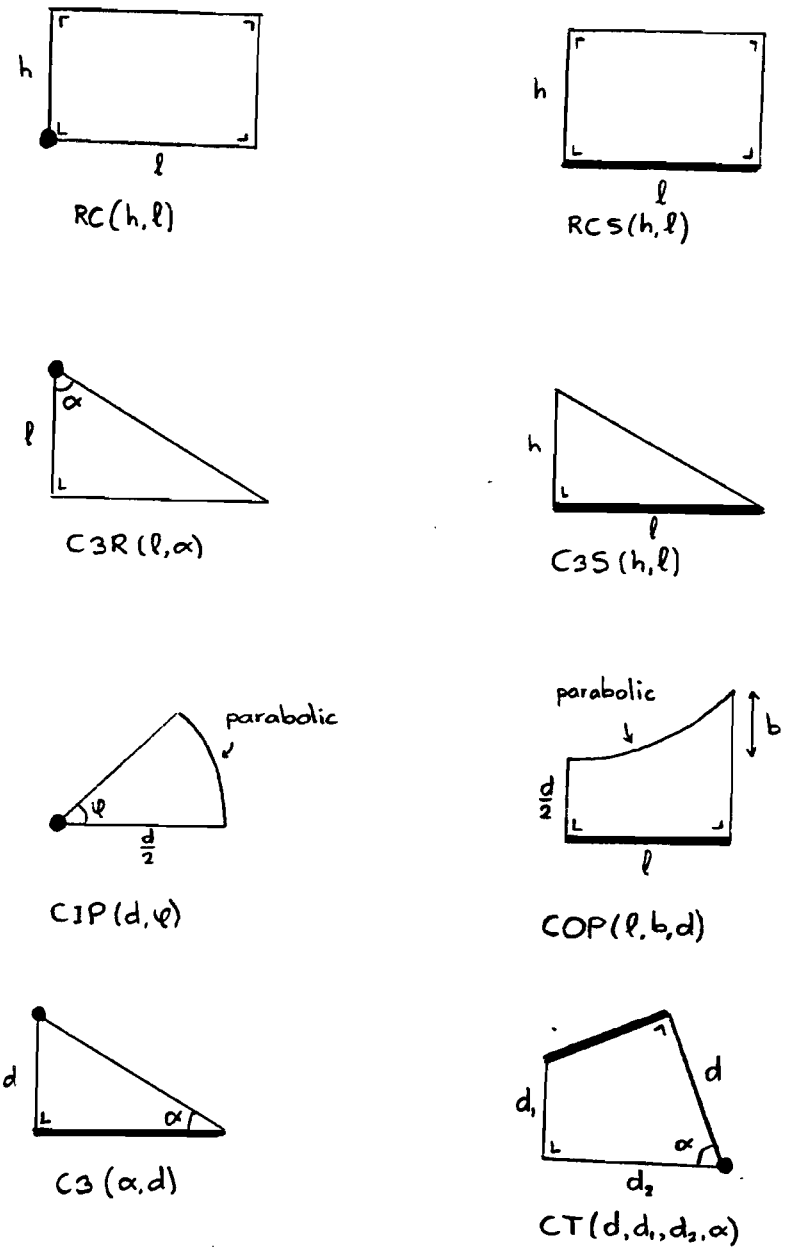



Figure 8: The regions for the cost functions.

- A square with a canal along one of the borders, with cost formula:

$$RCS(h,l) = wc_t hl\frac{h}{2}$$

(in fact, this formula was used in all previous designs).

- A right-angled triangle with a "source" in one of the corners, with cost function:

$$C3R(l,\alpha) = wc_t l^3\left(\frac{\tan\alpha}{6\cos\alpha} + \frac{1}{6}\ln\tan(\frac{\pi}{4}+\frac{\alpha}{2})\right).$$

- A right-angled triangle with a canal along one of its sides, with cost function:

$$C3S(h,l) = wc_t hl\frac{h}{6}.$$

- A triangle with a "source" in one of the corners and a canal along one of the sides. For its cost formula, we need to define some distances:

$$\begin{aligned} l_1 &= -d\tan\alpha + d\sqrt{\tan^2\alpha+1}, \\ l_2 &= \frac{d}{\tan\alpha} - l_1, \\ l_3 &= \frac{d}{2} - l_1\tan\alpha. \end{aligned}$$

Now the cost function can be expressed in three other cost functions, namely the function for a triangle with a canal along one of its sides ($C3S$) and two functions for areas with a parabolic border ($CIP$ and $COP$). These will be given later in this section. The costs are:

$$C3S(l_3+\frac{d}{2},l_2) + COP(l_1,l_3,d) + CIP(d,\frac{\pi}{2}-\alpha).$$

- A quadrangle with a canal along one of the sides and a "source" in one of the corners. Here we also have to define some variables:

$$\begin{aligned} l_1 &= \sqrt{d_1^2 + d_2^2 - d^2}, \\ l_2 &= \frac{1}{2}\sqrt{d_1^2 + d_2^2}, \\ \beta &= \arcsin \frac{l_1}{2l_2}. \end{aligned}$$

We have to consider four different cases:

- If $2\beta < \alpha$ and $d_2 > d_1$, the cost function is:

$$\begin{aligned} CT(d, d_1, d_2, \alpha) &= 2C3R(l_2, \alpha - \beta) - 2C3R(l_2, \beta) + C3R(d_1, \frac{\pi}{2} - 2\alpha + 2\beta) \\ &\quad + COP(l_1, \frac{l_2}{\cos\beta} - \frac{d}{2}, d) + CIP(d, 2\beta) \end{aligned}$$

- If $2\beta < \alpha$, $d_2 \leq d_1$, the cost function is:

$$\begin{aligned} CT(d, d_1, d_2, \alpha) &= 2C3R(l_2, \frac{\pi}{2} - \alpha + \beta) - 2C3R(l_2, \beta) + C3R(d_2, -\frac{\pi}{2} + 2\alpha - 2\beta) \\ &\quad + COP(l_1, \frac{l_2}{\cos\beta} - \frac{d}{2}, d) + CIP(d, 2\beta) \end{aligned}$$

- If $2\beta \geq \alpha$, $\alpha < \frac{\pi}{2}$, the cost function is:

$$CT(d, d_1, d_2, \alpha) = C3(\frac{\pi}{2} - \alpha, d) - C3S(d_1 \sin\alpha, \frac{d_1}{\cos\alpha})$$

- If $2\beta \geq \alpha$, $\alpha \geq \frac{\pi}{2}$, the cost function is:

$$CT(d, d_1, d_2, \alpha) = CIP(d, \frac{\pi}{2}) + COP(d, \frac{d}{2}, d) + RCS(d, d_2 - d)$$

The two cost functions $COP$ and $CIP$ still have to be defined:

$$\begin{aligned} COP(l,b,d) &= wc_t(\frac{lb^2}{2} - 2\sqrt{2bd}\frac{b^2}{5}) + RCS(\frac{d}{2}, l) \\ CIP(d,\phi) &= wc_t \tan(\frac{\phi}{2})\frac{d^3}{3}(\frac{1}{20\cos^4(\frac{\phi}{2})} + \frac{1}{15\cos^2(\frac{\phi}{2})} + \frac{1}{30}) \end{aligned}$$

With the aid of these cost functions, the total transportation costs for the area can be computed.

# CREATING AN OPTIMAL LAY-OUT FOR A PARKING LOT

# Abstract

In this report we show how we came to an optimal lay-out for a 70 x 35 metre corner parking lot.

We consider three ways of parking a car:

- right angle parking with the cars aligned side by side
- angle parking with the cars aligned side by side but slanted
- line parking: right angle parking with the cars aligned back to front

We show that the first option uses the best available space. This means that you start filling a parking lot with this type of parking spaces if you want to find an optimal solution.

We also consider the possibility of having more than one type of parking spaces in the lay-out, but in the end the best solution turns out to be a single type one: only straight angle parking with the cars aligned side by side is in it. In this particular lay-out, which is shown on page 17, we can locate 140 parking spaces on the parking lot.

After we got this result, we looked at the possibility of having an expert driver parking the cars. This man (or woman) would probably need less space to park the cars than the average driver and thus we might be able to locate more parking spaces. However, in our case this does not work. The optimal lay-out is different (see page 18), but the number of parking spaces is still 140.

# Contents

# 1 Introduction

The owner of a paved, 35 by 70 m, corner parking lot in a New England town hires someone to design the layout, that is, to design how the "lines are to be painted". To maximize the revenue for the owner it would be wise to squeeze as many cars as possible into the lot. However, inexperienced drivers may then have difficulties parking their cars, which can give rise to expensive insurance claims. To reduce the likelyhood of damage to parked vehicles, the owner might then have to hire expert drivers for "valet parking". On the other hand, most drivers seem to have little difficulty in parking in one attempt if there is a large enough "turning radius" from the access lane. Of course, the wider the access lane, the fewer cars can be accomodated in the lot, leading to less revenue for the parking lot owner.

Taking the above statement into account, two questions arise:

1. Which kind of layout leads to the maximum number of parking places in the lot.

2. Is it reasonable for the owner to hire expert drivers?

# 2 Different layouts

To handle the problem we make the following assumptions:

- The size of a car is limited by 5 m in length and 2.1 m in width which implies that no trucks or buses are allowed on the lot.
- The turning radius of a car, i.e., the radius of the smallest outer circle that a car can go around, is less than or equal to 5 m.
- On each side of a car 0.2 m are reserved to open the doors and get out of the car. (If the cars are aligned side by side, the space between two cars amounts to 0.4 m.)
- We consider three different types of parking a car:
  1. right-angle parking
  2. angle parking
  3. line parking

Furthermore, we first assume that the parking lot has one fixed entrance and one fixed exit which are connected by a one-way system.

## 2.1 Right angle parking

The first lay-out that comes to mind when thinking of parking cars on a parking lot is the right angle parking, with the cars aligned side by side.

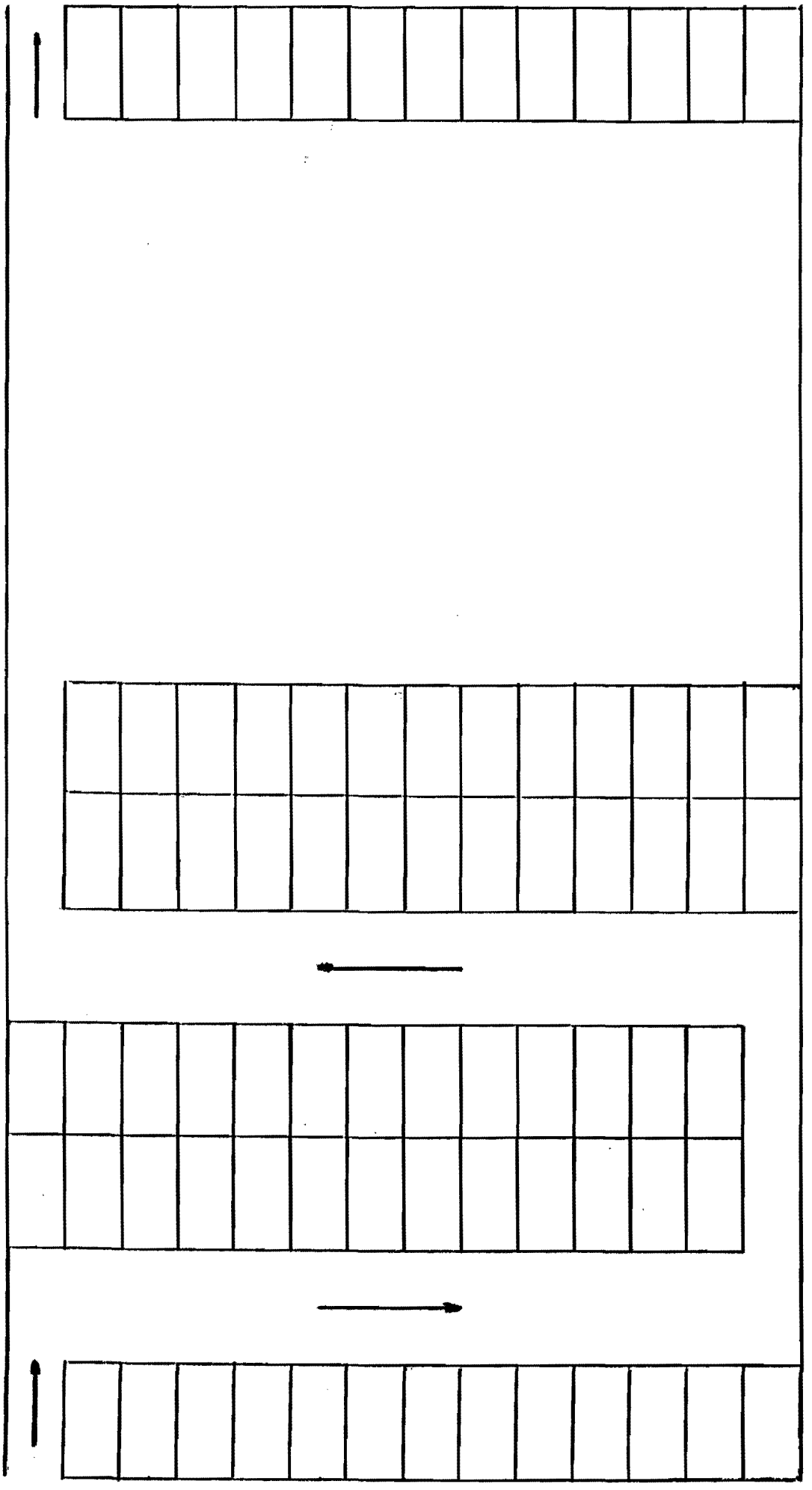
After deciding to concentrate on this type of parking, we need some dimensions: the size of the parking spaces and the width of the access lanes. We have already seen that we assume a parking space of 5 by 2.5 metres is enough to park any car and to be able to get out as well. Now what about the acces lanes?

We observed that the space needed to park a car in one attempt is the space occupied by a quarter of a circle with a radius equal to the car's turning radius. Since we assume that this radius is 5 metres, we need a square of 5 x 5 metres to park a car in one attempt. Consequently, the access lanes need to be 5 metres wide.

After we came this far, we sketched a possible lay-out, mainly to get an idea of how the parking lot would look like and also to get an idea on how many

parking spaces could be located. We arranged the parking spaces in blocks and let the access lane wind through these blocks. In this initial stage we used, as already mentioned, only one entrance and one exit.

The resulting lay-out is outlined on the next page. In this way we were able to locate 117 parking spaces. This lay-out has the advantage that cars can drive through the complete parking lot without changing directions, which makes traffic flow smooth.

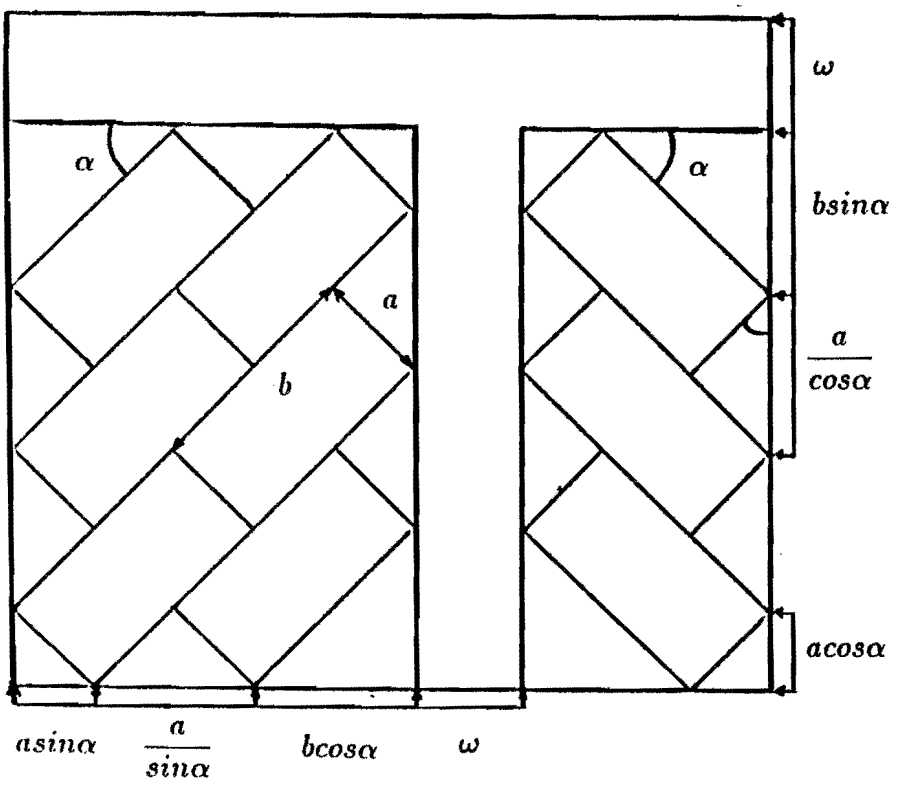


## 2.2 The angle parking

Let us now consider the angle parking. The advantage of this kind of parking is that the access lanes do not need to be as wide as in the case of a right angle parking. Therefore we are able to display more lanes along the x axis. But it is obvious that along the y axis less parking lots will be fitted.

Let us define a certain angle $\alpha$ such that $0 < \alpha < \frac{\pi}{2}$.

Fig 2.1:




Here $\omega$ denotes the width of the access lane.

We can see that if $\alpha$ decreases from $\frac{\pi}{4}$ to 0 then more parking lots may be fitted along the y axis. If we let $\alpha$ increases from $\frac{\pi}{4}$ to $\frac{\pi}{2}$ then more parking lots may be displayed along the x axis.

The disavantage of this layout is the waste of space caused by this angle $\alpha$. This is shown by the remaining "triangle" at the end of each parking place (cf Fig 2.1).

Thus the angled parking is more optimal than the right angle parking if and only if the area of its access lanes ($A_l$) and the "triangles"($A_{tr}$) is less than the area needed for the access lanes in the right angle configuration ($R_l$).

$$A_l + A_{tr} \leq R_l$$

Using the previous assumptions and the data of Fig 2.1 the layout of parking places must satisfy the following conditions:

a) Along the y axis:

$$b\sin\alpha + (K_1 - 1)\frac{a}{\cos\alpha} + a\cos\alpha + \omega \leq 35 \qquad (1)$$

where $K_1$ is the number of parking lots in a single lane, and $\omega$ is the width of the access lane.

b) Along the x axis:

$$2(a\sin\alpha + b\cos\alpha) + K_2(a\sin\alpha + b\cos\alpha + \frac{a}{\sin\alpha}) + \omega(K_2 + 1) \leq 70 \qquad (2)$$

where $K_2$ is the number of double parking lanes.

Therefore the total number of cars in the car park is:

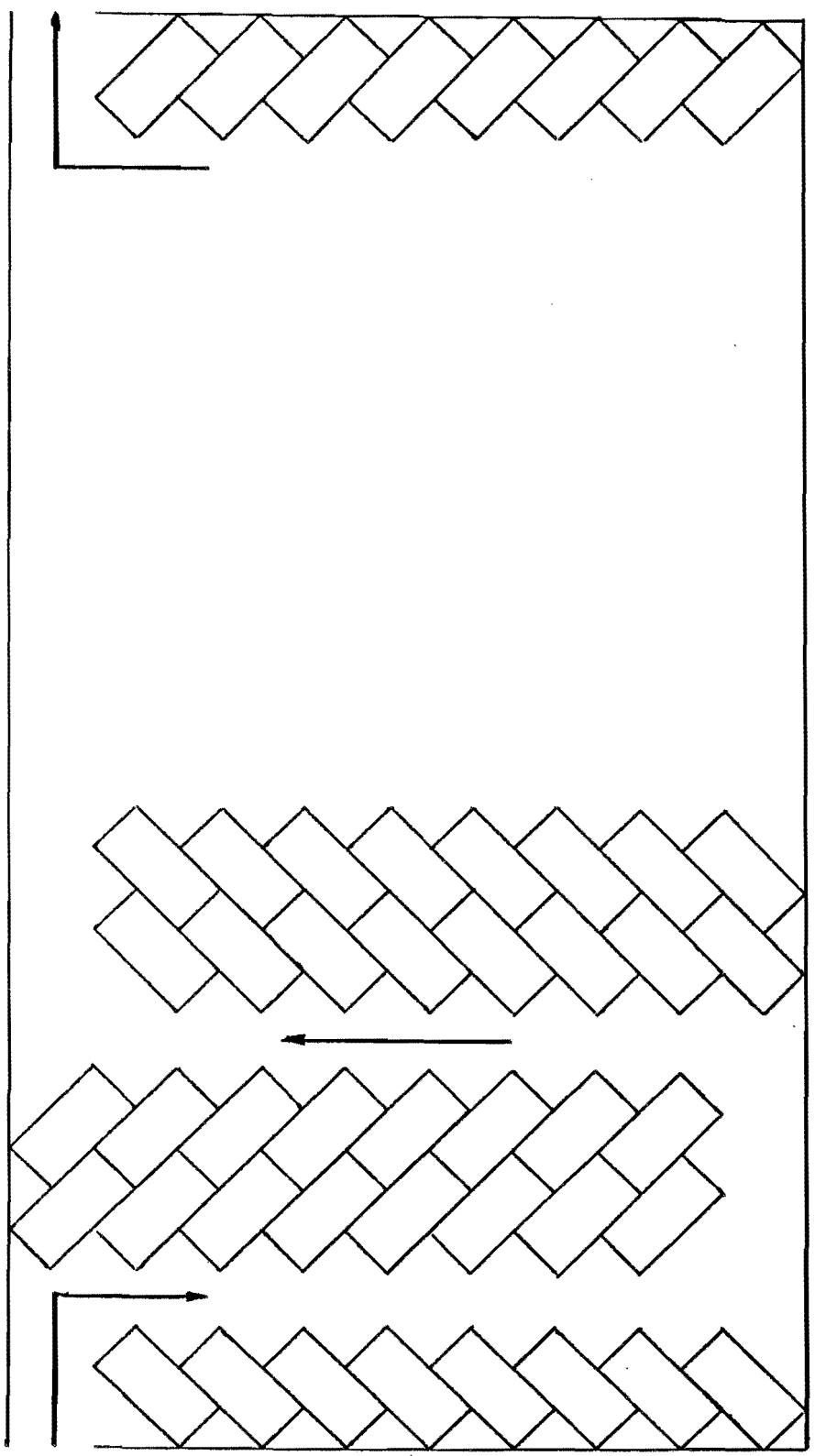
$$T_c = 2K_1(1 + K_2)$$

The following table illustrates the number of cars that may be fitted for a given $\alpha$ satisfying the equations (1) and (2):

| $\alpha$ | $K_1$ | $K_2$ | $T_c$ |
|---|---|---|---|
| $\frac{\pi}{3}$ | 6 | 6 | 84 |
| $\frac{\pi}{4}$ | 8 | 5 | 96 |
| $\frac{\pi}{6}$ | 10 | 4 | 100 |

The next picture (Fig 2.2) illustrates the optimal layout for $\alpha = \frac{\pi}{4}$.

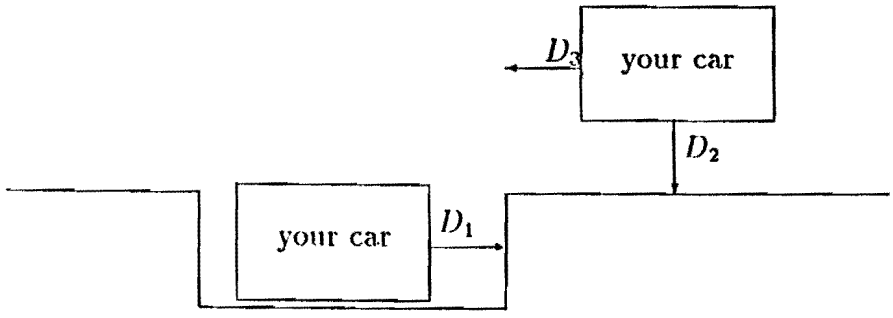
**Fig 2.2:**



## 2.3 The line parking

Finally we want to establish how many cars can be fitted into the parking lot, if one car is parked behind another, which we call "line parking". First of all we determine the space needed to park a car in such a way. We follow [1].

**Fig 3.1:**




Drivers should know that the best way of parking a car behind another is to reverse. To get your car parked in one attempt you should start the manoeuvre in the position determined by the distances $D_2$ and $D_3$ (see the following table) You end up with distance $D_1$ remoted from the car in front of yours. Therefore the length of a parking place is equal to the sum of the length of your car and $D_1$, whereas $D_2$ enters into the determination of the width of the access lanes. The table displays these crucial quantities for a sample of common cars.

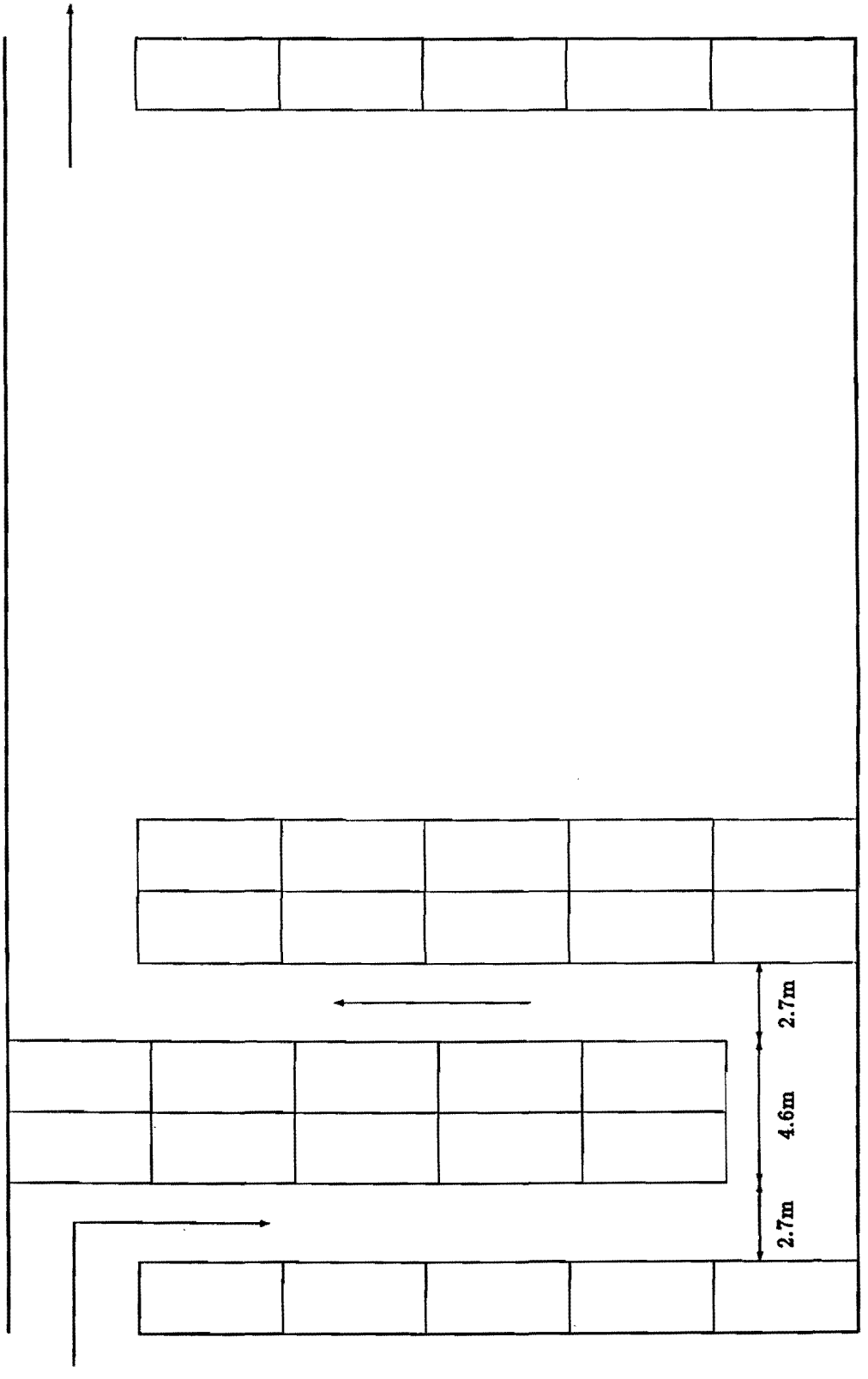
| model | length $+D_1$ | $D_2$ | $D_3$ |
|---|---|---|---|
| Volvo 440 | 5.84 | 0.09 | 0.31 |
| Volvo 940 | 5.85 | 0.01 | -0.38 |
| Ford Escord | 5.72 | **0.16** | 0.56 |
| Ford Sierra | 5.66 | 0.03 | -0.05 |
| Opel Astra | 5.45 | 0.10 | 0.44 |
| Opel Omega | **6.20** | 0.04 | 0.05 |
| VW Golf | 5.68 | 0.15 | 0.05 |
| VW Passat | 5.87 | 0.03 | 0.00 |

Obviously the size of a parking place has to be 6.2 by 2.3 m (to get out of the car on the driver's opposite side use the access lane!). The width of the access lanes has to be at least 2.26 m. However, drivers should be able to turn around at the end of a parking row without going back and forward. Taking

into account that the turning radius of a car is at most 5 m, the width of two access lanes and two rows of parked cars has to be at least 10 m which implies access lanes of width 2.7 m.

Fig 3.2 illustrates the layout for a line parking. We are able to locate 95 cars on the parking lot. Changing the direction of the parked cars (i.e.,the cars are parallel to the x-axis) leads to 99 parking places.

**Fig 3.2:**



# 3 An optimal solution

In the previous chapter we described three kind of parking. We also considered a one system inside the car park to be able to access any parking place. This involves a lot of space area needed for the access lane. However to maximise the number of parking places is equivalent to minimise the area of the access lane. In order to do so, let us make an important assumption.

As stated in the original problem we consider a corner parking with dimension 70 by 35m. Let us assume that a road is around two perpendicular sides of the car park (cf Fig 4.1). Then the ways out and ways in, are possible along this two sides.

As has been shown before, the right angle parking would be the most optimal one in this kind of situation. Therefore we now may try to display different layouts. This is illustrated by the following figures. In the first layout considered (Fig 4.1), a problem occurs. We covered 55 by 35m of the car park area, but in the remaining area (ie 15 by 35m) it might be wise to look at the other possible configurations individually.

a) Right angle parking (Fig 4.2):

In this case it is possible to fit 28 cars, which represents 66 percent of the actual used space.

b) Angle parking (Fig 4.3):

To find the optimal number of cars in this configuration we need to consider similar inequalities to equations (1) and (2). This yields:

$$2b\cos\alpha + 2a\sin\alpha + \frac{a}{\sin\alpha} + \omega \leq 15 \tag{3}$$

$$b\sin\alpha + (K_1 - 1)\frac{a}{\sin\alpha} + a\cos\alpha \leq 35 \tag{4}$$

Thus we have $K_1 = 7$ and $\alpha = 58°$. So in this case 21 cars can be parked, using 50 percent of the area.

c) Line parking (Fig 4.4):

Here 20 cars could be fitted, using 47.6 percent of the area.

**Fig 4.1:**

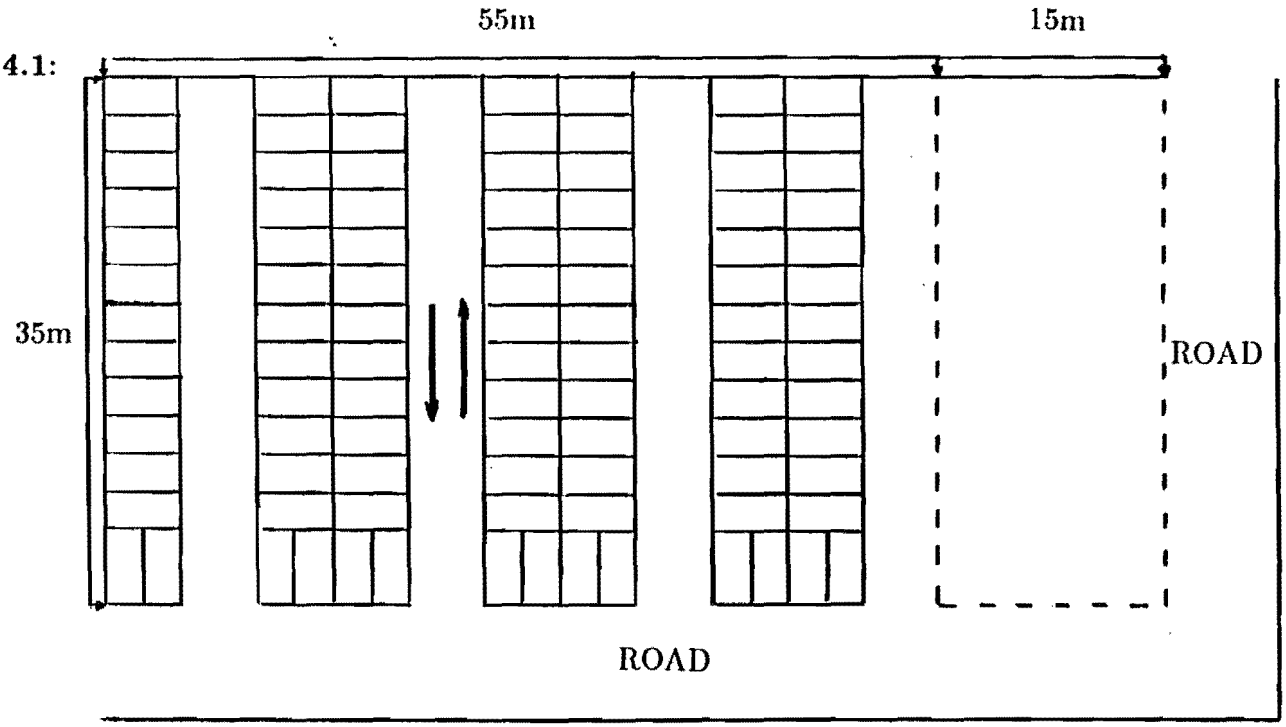



**Fig 4.2:**

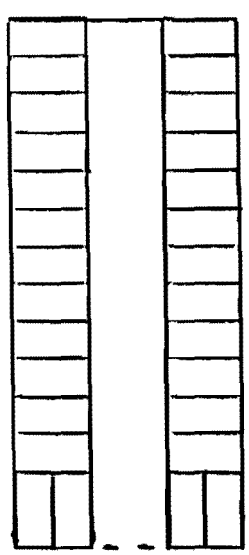


**Fig 4.3:**

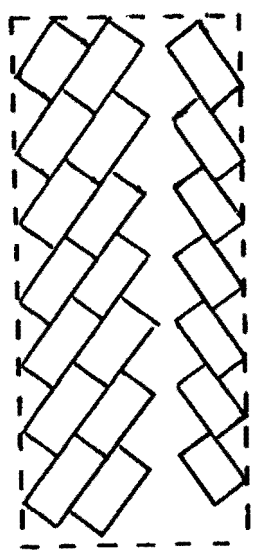
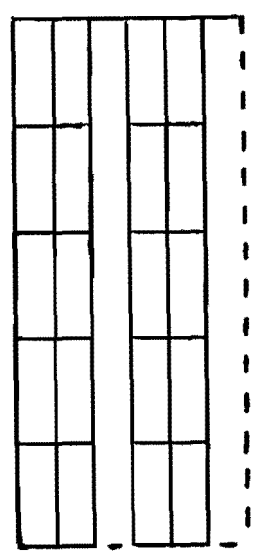


**Fig 4.4:**



Thus it is obvious once again that the right angle parking is the optimal display. Then matching Fig 4.2 with Fig 4.1 yields Fig 4.5 where 126 parking places are available.

But yet the space used by the access lanes seems not to be a minimum. Instead of displaying the access lanes along the **y** axis, we could display them along the **x** axis. This leads us to Fig 4.6, where 140 spaces are available, which represents 71.4 percent of the area used. In that respect it appears to be the optimal solution. However to have a convenient layout it is possible to remove the 4 spaces standing at the far left hand side of the middle row, in order to be able to drive around easily (Fig 4.7). In this case 136 cars can be parked. Considering that the car park would be full with 196 cars, this layout is a convenient optimal solution.

Fig 4.5:

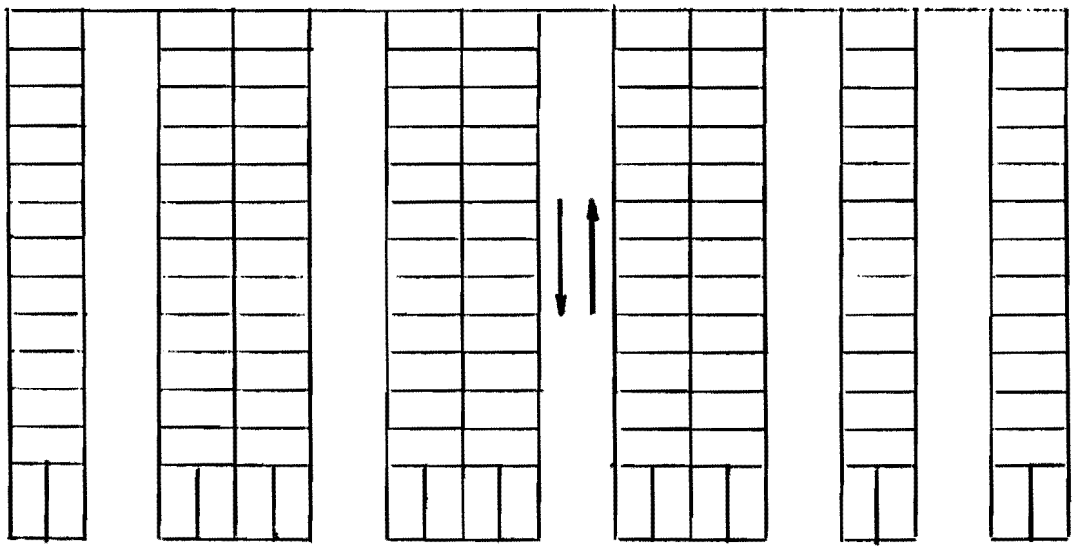


Fig 4.6:

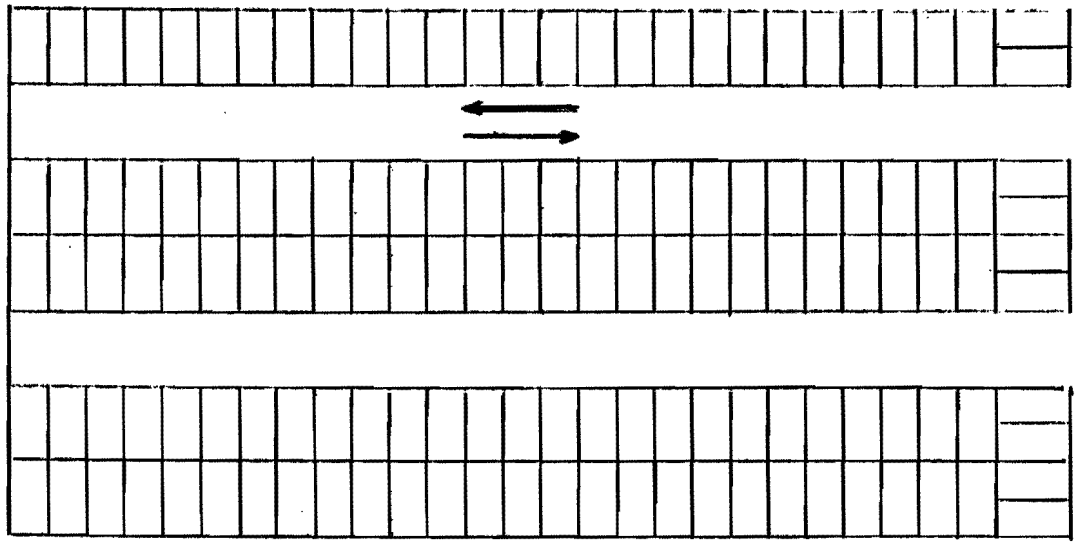
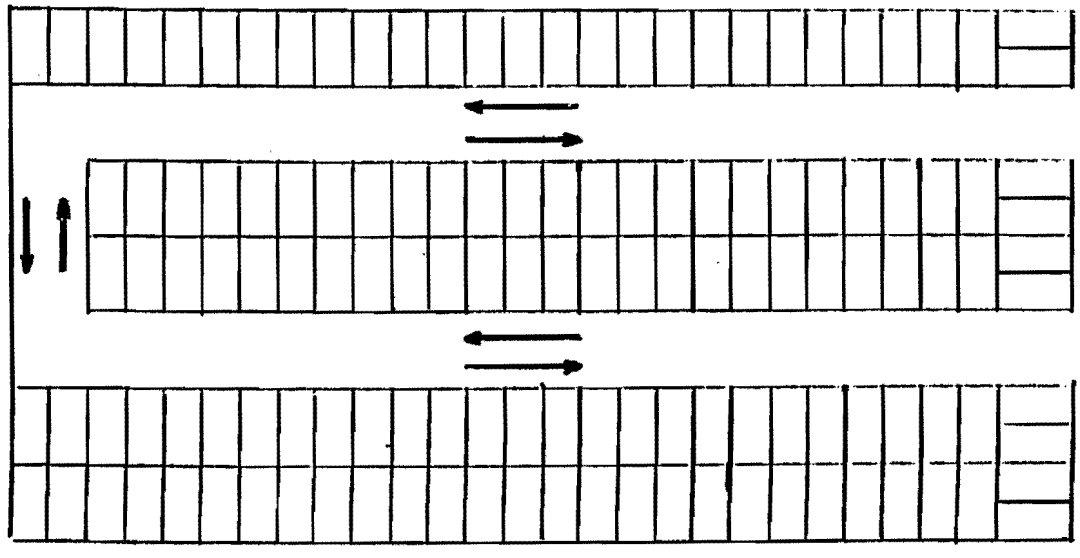


Fig 4.7:



# 4 Using an expert driver to park the cars

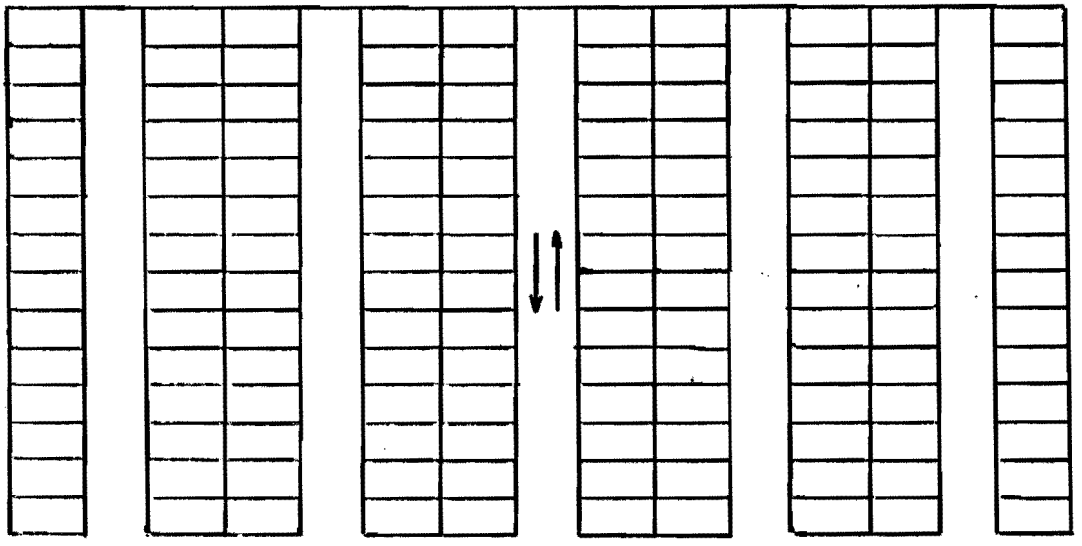
What do we gain from using an expert driver to park the cars? Note that we only look at **available parking space** (i.e. space left of parking lot after putting access lanes in). We do not look at economical or other benefits. For instance: having an expert driver around, makes it easier to ask higher parking fees because you can present your parking lot as being guarded.

In case of the right angle parking there is an advantage: access lanes can be smaller, since the restriction that cars must be parked 'in one go' does not hold anymore. In both the angle parking and the line parking there is no difference.

Since we have an optimum solution for the average driver with only right angle parking spaces, we do look at the expert driver case. Knowing that we can park 140 cars in the first case, we expect to be able to raise that figure.

As already stated: the expert driver needs less wide access lanes. We assume a 20 % reduction, which leads to 4 metre wide access lanes. From Section 3 we know two good solutions (cf Fig 4.5 and 4.6), where the first has its access lanes along the **y** axis and the second one, yielding the optimal solution of 140 places, has them along the **x** axis. We can see that the optimal one does not benefit from the reduction of the access lanes: we get a 2 by 70 m area extra for parking spaces, but we can not use it, since a parking space is 2.5 metres wide. On the other hand, from Fig 4.5 we see that we get an extra 5 by 35 m area for parking spaces, which is exactly enough to fill in an extra block. This extra block of parking spaces can be reached without changing the original access lane lay-out.

This optimal lay-out, as shown in Fig 5.1, has a capacity of 140 cars, with a 71.4 % usage of the total available space for parking spaces, the same figures as those for the optimal lay-out for the average driver, shown in Fig. 4.6. Consequently, the choice between hiring an expert driver or not will, in this case, depend only on economical or other benefits.



# 5 Conclusions

In this report we developed two lay-outs we consider being optimal. The first one in case the average driver must be able to park his/her car in one attempt, the second in case an expert driver is hired. Both solutions yield 140 parking spaces, although to get a more 'costumer friendly' lay-out, the average driver case reduces to 136 parking spaces (see Section 3 and Fig 4.7).

Even then, there is no (big) difference between the two solutions, if you look at them from a mathematical viewpoint. The decision whether to hire an expert driver or not is therefore an economical one.

Even if not economically beneficial at first sight, it can be a good idea to hire an expert driver when car thefts or car burglaries are a problem in the area. A guarded parking lot will then attract more people.

As far as other parking lots are concerned: there will allways be the need for adjustments to their particular shapes, but the principle stays the same.

- look for the optimal way of parking a car,
- fill the parking lot 'as far as possible',
- construct solutions from all types of parking for the remainder,
- and finally: fit the two pieces together.

# References

[1] A. den Boer, R.J.H. du Croo de Jongh, S.R. Tiourine. *Parking a car in the smallest possible space.* Modelling report, Univ. Eindhoven, April 1992.

# EMERGENCY FACILITIES LOCATION

# Contents

# 1 The problem

## 1.1 Introduction

This project concerns the township Rio Rancho. Here, all kinds of accidents happen, and since Rio Rancho hitherto did not have its own facilities, emergency services had to come over from neighbouring cities. The accident-rate increased through the years and this became an impractical situation. After some years of saving, Rio Rancho now has secured funds to erect two emergency facilities in 1993, each of which will combine ambulance, fire and police services. The task is to locate the two facilities in an optimal way.

## 1.2 Data

Figure 1 presents a simplified city map of Rio Rancho. The horizontal and vertical lines indicate streets that divide the city into blocks and the numbers correspond the number of emergencies per square block for 1992. The L region in the north is an obstacle, while the rectangle in the south is a park with a shallow pond. It takes an emergency vehicle an average of 15 seconds to go one block in the North-South direction and 20 seconds to move one block in the East-West direction.

| | | | | |
|---|---|---|---|---|
| 20 | 34 | 28 | 23 | 14 |
| 33 | 32 | 28 | 21 | 12 |
| 24 | | 23 | 16 | 11 |
| 20 | | | 16 | 12 |
| 19 | 18 | 29 | 18 | 16 |
| 16 | 22 | 34 | 27 | 22 |
| 11 | 18 | 33 | 15 | 12 |
| 17 | 10 | | 30 | 25 |
| 27 | 27 | | 25 | 18 |
| 31 | 23 | 12 | 31 | 27 |

⇑ North

Figure 1: Demand in Rio Rancho in 1992

## 1.3 Problem formulation

The task is to optimally locate the two facilities, in the two following cases:

- The demand is concentrated at the center of the blocks and the facilities are located on corners.
- The demand is uniformly distributed on the four streets bordering each block and the facilities may be located anywhere on the streets.

Since both problems are almost equivalent (this will be shown in Section 3), we decided to solve additionally the following problem:

- How many ambulances and fire-engines should both emergency centres keep, such that the number of times an accident occurs while there is no free vehicle, is acceptably low?

# 2 Where to locate the emergency centres

## 2.1 Introduction

The first problem is to optimally locate the two emergency centres. The demand for facilities is concentrated at the centre of each block and the facilities can be located on corners.

## 2.2 Assumptions

We make the following assumptions:

- Every accident is served by the emergency centre that is closest to the location of the accident. If an accident takes place on equal distances of both emergency centres, it is served by either of them with probability 0.5.
- There are enough vehicles present to drive to all emergencies.
- Every vehicle drives along the shortest route (in time), given by the map.
- A vehicle has reached the emergency if it has reached one of the four corners of the block in which the accident happened, as shown in Figure 2.

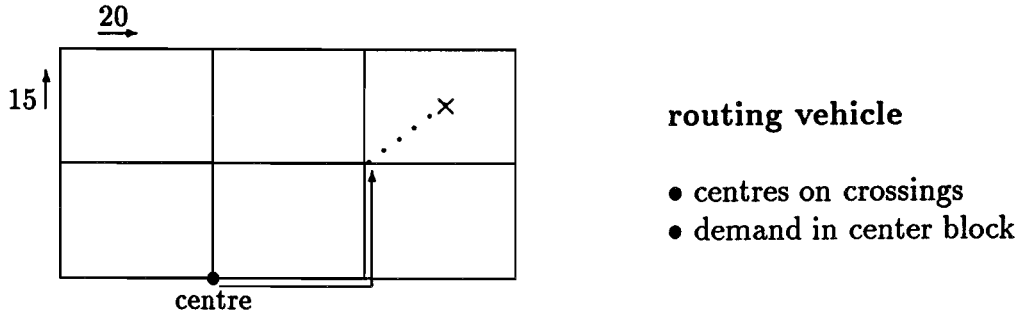



Figure 2: Route from centre to accident

- An emergency centre is located optimally if the total number of driving hours per year of its vehicles from the centre to the different emergencies is minimal.

One remark concerning the last assumption:
Another criterion to define optimality is minimizing the maximal driving time to an emergency. One can imagine that for a fire, it is essential to be on the place of the accident within a certain time, the whole town might burn down. When the total number of driving hours is taken as a criterion, an emergency centre will probably not be located in the neighbourhood of a 'quiet' area, because the weight of this area on the total driving time is small. Therefore it will take a long time to reach this area, maybe too long. Minimizing the maximal driving time will take care of this problem. We felt however that because the driving times are rather small in this problem (one can cross the town within 250 seconds), we could not let many people suffer for the sake of one single person and gain only a few seconds. Hence we choose to minimize the total driving time and do not go into minimizing the maximal driving time.

## 2.3 Writing a program to compare driving times

We solved the problem in a very straightforward way.

First we numbered all street corners from 0 to 65, starting from the lower left hand corner (South-West) and going to the right (South-East), then starting from West going East again, but one street up North, etc. The demands for facilities are given in a $5 \times 10$ matrix 'Demand'.

Then we wrote a computer program that has the following structure: (in Pseudo-Pascal)

```
BEGIN
    MinimalDrivingTime := MAXINT;
    FOR i = 0 TO 65 DO                          {Location of first centre}
    FOR j = i + 1 TO 65 DO                      {Location of second centre}
    BEGIN
        TotalDrivingTime := 0;
        FOR k = 1 TO 5 DO                       {First coordinate of demand}
        FOR l = 1 TO 10 DO                      {Second coordinate of demand}
        BEGIN
            D1 := DrivingDistance(k , l , i);
            D2 := DrivingDistance(k , l , j);
            IF D1 < D2 OR ((D1 = D2) AND (RANDOM(0,1) < 0.5))
            THEN TotalDrivingTime :=
                    TotalDrivingTime + D1*Demand(k , l)
            ELSE TotalDrivingTime :=
```

```
                TotalDrivingTime + D2*Demand(k , l);
        END;
        IF TotalDrivingTime < MinimalDrivingTime
        THEN
            BEGIN
                MinimalDrivingTime := TotalDrivingTime;
                LocationFirstCentre := i;
                LocationSecondCentre := j
            END;
    END;
    MinimalDrivingTime := 2*MinimalDrivingTime;
END.
```

In words, for every possible pair of locations of the centres, the one is picked that gives the smallest total driving time. This total driving time is computed in the following way: For every city block, the distance to the nearest by centre is computed, and is multiplied by the total demand in the block. Twice the sum of all these numbers equals the wanted value.

## 2.4 Results

Implementation of the program gives the following results: The first emergency centre can best be located on position 21, the second emergency centre can best be located on position 50. The total two-way driving time per year then is 15.04 hours. The optimal emergency centre locations are shown in Figure 3.

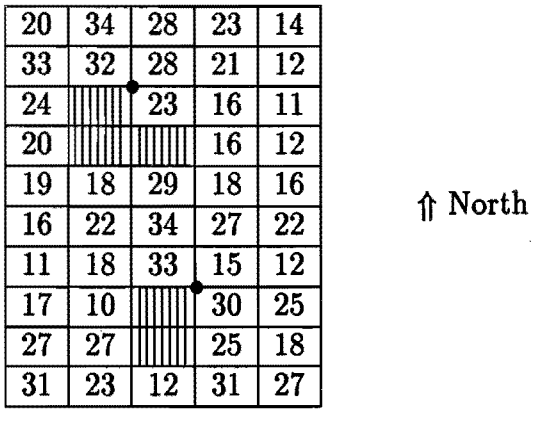



Figure 3: Optimal location of two emergency centres in Rio Rancho

The next figure shows which block has been assigned to which emergency centre.

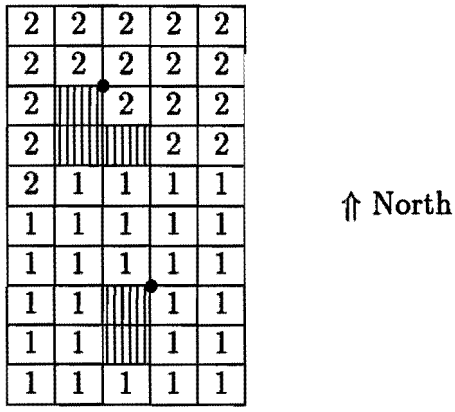



Figure 4: Assignment of city blocks to emergency centres

# 3 Alternative locations

## 3.1 Introduction

In the previous chapter, facilities were placed on the corners on the street while accidents were supposed to happen in the middle of the blocks. Instead of this rather easy model, we will now consider a more continous emergency problem: accidents happen anywhere on the street according to an uniform distribution and the facilities are placed also anywhere on the street.

## 3.2 Assumptions

We make the same basic assumptions as in paragraph 2.2. That means, we have the same optimization procedure, vehicles driving along the shortest route and coming from the nearest centre.

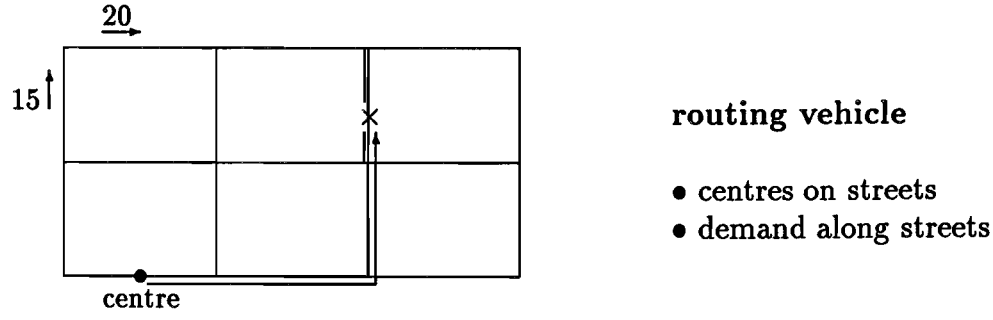



Figure 5: Route from centre to accident

The calculated distance and distribution of accidents however are different:

- Because the accidents can take place anywhere in the street, the car now has to drive to the specific place in the street instead of to the nearest corner. This is shown in Figure 5.
- The emergencies happen somewhere in the streets according to a uniform distribution. The total demand in a street is given by the demand in the

adjacent blocks (see Figure 1). For example, with demand 20 and 34 in the blocks along the street, the demand in the street is $(20+34)/4 = 13.5$ per year.

## 3.3 Expected driving time

The fact that emergencies now occur in the streets instead of in the centre of the blocks has large implications for calculating the driving time. Let us first consider the assignment of emergencies to a centre. In the previous model, the accident happened in the centre of a block, and the block was assigned to the nearest centre. With locations on the whole street, however, some streets are assigned to one centre and some are divided into two parts assigned to different centres.

The total driving time in which we are interested is now a stochastic variable, equal to the weighted sum of the weighted expected driving times to all streets (having made the assignments). The weighted driving time is the demand of the street times the integral of the distance to the emergency centre.

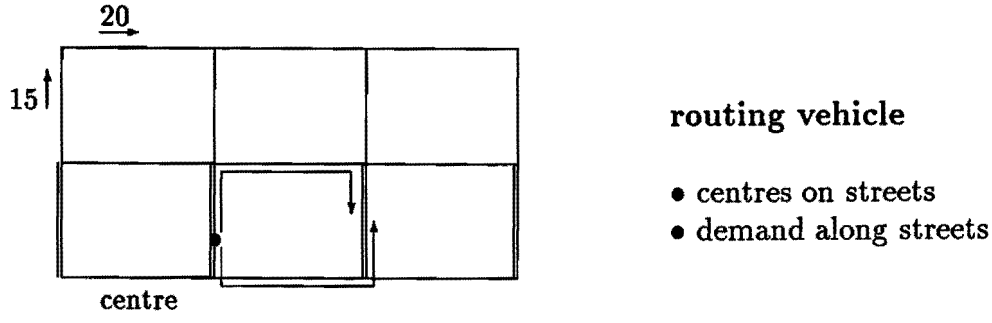



Figure 6: Separated routes

In general, for a street assigned to a centre there will be a fixed route from centre to accident, independent from the exact place of occurence. In Figure 6 we see, however, that for certain streets there may be two routes to the accident, depending upon whether the 'upper' or 'lower' variant is quicker. The streets for which two routes exist are easily found. They lie parallel to the street of the emergency centre in a configuration similar to a ladder (the double-lined streets, see Figure 6).

Now we have three types of streets:

- Streets that are assigned partially to both centres,
- Those situated on the ladder, with two different routes,
- Regular streets, with just one assigned nearest centre and one route.

Now for all streets the integrals of distance can be reduced. First the streets of type one and two are split. Then, all integrals can be replaced by the distance from the assigned centre to the middle of the (split) street.

The model in which the emergencies take place anywhere in the streets can be reduced to a model in which accidents occur in fixed points. The model therefore is like in the previous section, with the exception that the centres may be placed everywhere in the street. The problem is easy when the total driving time of an observed location of centres is lineairly depending on the distances from centre to streetcorner. Then the optimal placement is on a corner. But now the driving time is a quadratic function of the distance to the corner, due to the splitting of streets with two different routes. That means an optimal location can be anywhere on the street.

## 3.4 Implementing the model

Until now we didn't discuss the interpretation of continuous locations and the way of choosing routes. Before continuing the implementation of the model, it is however very useful to see for reality of the emergency serving.

After a call for help, one has to decide which centre should deliver a vehicle. The choice definitely depends on the street where the accident happened, but sometimes also on the place in the street. However, to ask for the exact distance between accident and corner seems far from realistic. When the centre nearest to the middle of the street is assigned, the distance will differ at most 20 seconds from the real nearest centre. For streets on the ladder the fixed route will be at most 10 seconds longer than the shortest route.

With those restricted deviations in time, a fixed route and a fixed assignment for all streets seems reasonable. We can then replace the accidents happening in the streets by accidents in the middle of the streets. Furthermore the total driving time depends linearly on the placement of the centres, and therefore an optimum is found on the corners. With some slight adjustments of the map

the problem can be reduced to the model of Section 2, so the implementation of this model is finished.

# 4 Determining the best number of vehicles

## 4.1 Introduction

Now that we have optimally located the emergency centres, we want to determine the minimal number of ambulances and fire-engines we have to keep, such that no people have to die without being served first.

## 4.2 Assumptions

Because too few data are provided for answering this question properly, we make the following additional assumptions. (cf. Section 2.2)

- Each emergency in the demand figures needs one vehicle.
- The demand per square block in Rio Rancho can be split into a demand for ambulances and a demand for fire-engines, according to Figures 4 and 5. These demands are independent.
- The demands for ambulances and fire engines are made according to independent Poisson processes (with possibly different parameters).
- When there is a demand but there is no free vehicle at the nearest emergency centre, the patient will not be helped and will probably die or burn.

The third assumption concerning the distribution of the demand for vehicles helps us to find a solution to the problem in a very easy way. A discussion of the legalness of this assumption can be found in Section 5.

| | | | | |
|---|---|---|---|---|
| 19 | 31 | 26 | 21 | 13 |
| 31 | 29 | 27 | 17 | 9 |
| 23 | | 19 | 11 | 3 |
| 13 | | | 7 | 1 |
| 7 | 5 | 8 | 1 | 3 |
| 1 | 3 | 5 | 3 | 7 |
| 3 | 7 | 9 | 3 | 5 |
| 11 | 1 | | 17 | 19 |
| 14 | 15 | | 12 | 12 |
| 18 | 12 | 5 | 19 | 11 |

⇑ North

Figure 7: Demand for ambulances in Rio Rancho in 1992

| | | | | |
|---|---|---|---|---|
| 1 | 3 | 2 | 2 | 1 |
| 2 | 3 | 1 | 4 | 3 |
| 1 | | 4 | 5 | 8 |
| 7 | | | 9 | 11 |
| 12 | 13 | 21 | 17 | 13 |
| 15 | 19 | 29 | 24 | 15 |
| 8 | 11 | 24 | 12 | 7 |
| 6 | 9 | | 13 | 6 |
| 13 | 12 | | 13 | 6 |
| 13 | 11 | 7 | 12 | 15 |

⇑ North

Figure 8: Demand for fire-engines in Rio Rancho in 1992

Since the accidents now are assumed to occur according to Poisson processes, the problem has become stochastical and therefore it is not possible to reduce to zero the probability that there is no free vehicle. What we can do is reduce this probability to a reasonably low value, but then we have to define what is reasonable. We decided to allow that

- At most once a year the need for a ambulance can not be fulfilled
- At most once every ten years the need for a fire-engine can not be fulfilled

It is not yet known for how long the vehicles will be needed when serving an accident. Such a serving time of an ambulance will include going into a house,

lay the patient on a stretcher, etc. For a fire this service time will include unrolling the hoses and putting out the fire, etc. We chose to use the following average service times.

- Apart from driving times, an ambulance is needed for 20 minutes per accident on average.
- Apart from driving times, a fire-engine is needed for 5 hours per accident on average.

## 4.3 Computing the number of vehicles needed

From queueing theory, we know that when clients arrive according to a Poisson process with parameter $\lambda$ at an office with $s$ servers while there is no waiting room, the probability $P_s$ that they have to leave because there is no server free is given by the following formula:

$$P_s = \frac{\frac{\rho^s}{s!}}{\sum_{l=0}^{s} \frac{\rho^l}{l!}} \tag{1}$$

Here, $\rho$ denotes the utilization factor, which is given by $\lambda$ times the average service time. $P_s$ is usually called the probability of blocking.

We can use this formula taking as the average service time the average driving time of an ambulance or fire engine added to the time that a vehicle is needed on the place of the accident. As these times are known, we only have to determine the parameters of the Poisson processes describing the occurrence of the accidents. We will use that the Poisson parameter describing service $j$ ($j \in$ {ambulance, fire engine}) for centre $i$ ($i \in \{1, 2\}$) equals the average demand per year for vehicles of service $j$ for centre $i$, which can be approximated by the total demand in 1992 for vehicles of service $j$ for centre $i$.

## 4.4 Results

Implementation of the procedure described in the previous section gives the following results:

Centre 1 (South), ambulances.

- Average demand per year $\lambda$ = total demand in 1992 = 364

- Average service time = 21.5 minutes
- Utilization factor $\rho = 0.016$
- Blocking probability when keeping 1 ambulance = 0.016 (this means once every 2 months)
- Blocking probability when keeping 2 ambulances = 0.00013 (this means once every 21 years)

Centre 1 (South), fire-engines

- Average demand per year $\lambda$ = total demand in 1992 = 229
- Average service time = 5.06 hours
- Utilization factor $\rho = 0.132$
- Blocking probability when keeping 1 fire-engine = 0.12 (this means once every 13 days)
- Blocking probability when keeping 2 fire-engines = 0.0077 (this means once every 7 months)
- Blocking probability when keeping 3 fire-engines = 0.00034 (this means once every 13 years)

Centre 2 (North), ambulances.

- Average demand per year $\lambda$ = total demand in 1992 = 79
- Average service time = 21.5 minutes
- Utilization factor $\rho = 0.0035$
- Blocking probability when keeping 1 ambulance = 0.0034 (this means once every 4 years)

Centre 2 (South), fire-engines

- Average demand per year $\lambda$ = total demand in 1992 = 307
- Average service time = 5.06 hours

- Utilization factor $\rho = 0.177$
- Blocking probability when keeping 1 fire-engine = 0.15 (this means once every 8 days)
- Blocking probability when keeping 2 fire-engines = 0.013 (this means once every 3 months)
- Blocking probability when keeping 3 fire-engines = 0.00077 (this means once every 4 years)
- Blocking probability when keeping 4 fire-engines = 0.000034 (this means once every 95 years)

## 4.5 Recommendated capacities

To meet with our safety restrictions, in the first centre at least 2 ambulances and 3 fire-engines are needed while in centre 2, at least 1 ambulance and 4 fire-engines are needed. Of course, Rio Rancho can reduce cost by taking less vehicles, and increase risk, or it can make the town safer than suggested by keeping more vehicles.

We come back to a remark made in Section 4, concerning the assumption that the demand for emergency vehicles is made according to Poisson processes.

In queueing theory the assumption of having Poisson arrivals is very popular. The distribution gives way to very nice computation methods and it is known that when arrivals occur completely at random, it even gives a very good approximation of the real distribution. Without making this assumption, we were probably not able to find a solution to our problem. What we want to know now is if we made a very big mistake by doing this.
Making the demand arrive according to independent Poisson processes implicitly implies the following:

- Accidents that need an ambulance as well as a fire engine do not occur.
- Accidents that need more of the same sort of vehicles at the same time do not occur.
- Accidents occur during the whole day and not for example only in peak hours.
- Accidents occur during the whole year at the same level.

All this implies that all accidents are about equally big and there are no season effects. Of course in real life this will not be the case.
We think that the only possibility to deal with this problem is to divide the four arriving processes into even more processes. New data about the numbers of accidents occuring per block in the year are needed, but this time divided into groups. There is one group of accidents that needed 1 ambulance, one group that needed 2, etc. Then with these new numbers, the method of Section 4 can be used again. This way more accurate solutions can be given.

# 5 Conclusions

The project first was stated as a location problem with optimisation of the driving times. We had two different models for emergency demand and possible places for the two centres, say a discrete and a continuous case. The two models appeared to lead to the same solution method however. Therefore we only optimised the discrete case, and made an extension with observing the needed capacities of the centres.

The optimal location of the centres in Rio Rancho is determined by full search. For all possible locations of the two centres, we computed the total driving time. This total time is the sum of the driving times to all accidents happened in 1992. Thus for each location, the distances from the accidents to the nearest centre are computed, and multiplied by the fixed speed of the emergency vehicles. Obstacles in the map are included. The optimal location is shown in Figure 3, Section 2.

The additional problem is to determine the needed capacity of the optimal placed centres. Therefore the emergency figures of 1992 are split into a demand for ambulances and a demand for fire-engines. Furthermore the average time to serve an accident is chosen, and the occurence of accidents is assumed to be a Poisson process. Then the probability of running out of vehicles has to satisfy some safety restriction. This gives us the need of ambulances and fire-engines of centre 1 and the need of centre 2.

The fact that a solution is found does not mean the problem is finished. Probably it will never be, but based on this model we already can make some remarks. The assumptions which simplify the problem can be replaced by more realistic assumptions in a refined model. This is mentioned in the previous sections, but here we state some general points for further research:

- Consider a more realistic map (curves, different types of roads, traffic lights, 'moving obstacles' like roads being repaired on different places consecutively).
- Make less simplifying assumptions about the arrival of emergencies in the additional problem. The Poisson parameters may be time dependent, and the number of needed vehicles may be more than one per accident.

# ICEBERGS FROM ANTARCTIC

## Abstract

The cost of desalinating sea water conventionally in the Persian Gulf is high and requires extensive amounts of oil. Scientists suggested that it could be less expensive to tow icebergs from the Antarctic and melt them to produce drinking water.

With a few assumptions we made differential equations from the given data and got a control problem with the velocity as control function. We solved the nonlinear optimization problem resulting from discretisizing the velocity function. In this way we found that this method is indeed cheaper. Ca. 10% of the costs could be saved in this way.

This result is heavily dependent on the form of the iceberg. So a final result can be given only after experimental determination of this dependency.

# Contents

# 1 Introduction

## 1.1 Problem Description

The cost of desalinating sea water conventionally in the Persian Gulf is high (viz. $10p/m^3$) and requires extensive amounts of oil. Some time ago scientists suggested that it could well prove both practically feasible and less expensive to tow icebergs from the Antarctic, a distance of about 9600 km. Although some of the ice would undoubtedly melt it was thought that a significant enough proportion of the iceberg should remain intact right to the Persian Gulf.

Our task was to evaluate a strategy to produce the cheapest water for the Persian Gulf by towing icebergs and to decide on its economic feasibility.

## 1.2 Given Data

A programme of work was carried out to evaluate the practical problems associated with such a proposal and to quantify the factors that were likely to influence the economics of such a venture. Amongst other factors was identified the variability in the rental costs of the different-sized towing vessels plus the maximum loads they are able to tow (see Table 1).

| Ship size | Small | Medium | Large |
|---|---|---|---|
| Daily rental [ £] | 4.00 | 6.20 | 8.00 |
| Max. load [$m^3$] | 500,000 | 1,000,000 | 10,000,000 |

Table 1: Towing vessel data

It was also found that the melting rate of the iceberg depends upon the towing speed and its distance from the South Pole, at least, up to 4000 km away. Table 2 summarises the data available to assess the rates at wich icebergs melt.

| Towing speed $[km/h]$ | Distance from Antarctic $[km]$ | 0 | 1000 | > 4000 |
|---|---|---|---|---|
| 1 | | 0 | 0.1 | 0.3 |
| 3 | | 0 | 0.15 | 0.45 |
| 5 | | 0 | 0.2 | 0.6 |

Table 2: Melting rates of icebergs $[m/day]$

Finally, the fuel cost was found to be heavily dependent on the towing speed and the size of the iceberg, though it was relatively independent of the size of the towing vessel. The available data relating to fuel costs is summarised in Table 3.

| Towing speed $[km/h]$ | Iceberg volume $[m^3]$ | $10^7$ | $10^6$ | $10^5$ |
|---|---|---|---|---|
| 1 | | 12.6 | 10.5 | 8.4 |
| 3 | | 16.2 | 13.5 | 10.8 |
| 5 | | 19.8 | 16.5 | 13.2 |

Table 3: Fuel costs [ $£/km$]

# 2 Modelling of the problem

## 2.1 Assumptions

Since we have only a few measure points for the driving from Antarctic to the Persian Gulf, it is not possible to describe the whole dependencies exactly. Therefore we make the first assumption :

**Assumption 1** *All dependencies are partly linear between the given measure points.*

For the fuel costs we take this assumption for the logarithm of the volume, because the three measure points then lay on a straight line for every speed.

Obviously, we have to take into account the costs for travelling from the Persian Gulf to the Antartic. However, there is nothing said about these costs. So we make two assumptions for this travelling :

**Assumption 2** *The driving speed of a towing vessel to the Antarctic is 30 km/h.*

We think this assumption is realistic, since the ships drive in this direction without any iceberg, and so they can drive faster then with an iceberg.

**Assumption 3** *The fuel cost for one kilometer without any iceberg and velocity 30 km/h is 2 £.*

This value must be much smaller than the driving costs with an iceberg, but since the speed is much higher, it can not be near zero. So 2 £ is a possible value.

Since the values for driving from Antarctic to the Persian Gulf are only given for velocities between 1 km/h and 5 km/h, we think, a ship can in practice only drive with speeds in this interval when towing an iceberg. This leads to our next assumption :

**Assumption 4** *The velocity for towing an iceberg lies between 1 km/h and 5 km/h.*

The melting rate is given in meters per day. This means that the iceberg melts with this rate uniformly from all sides. When considering this unit as cubic meters per square meters per day, we need a relation between the surface and the volume of the iceberg. We considered several mathematical bodies and found, that for every body there is a constant factor $c$, such that the surface $S$ satisfies

$$S = cV^{\frac{2}{3}} \tag{1}$$

This factor depends on the form of the body :

| | |
|---|---|
| ball | $c = 4.6$ |
| cube | $c = 6$ |
| rectangular solid | $c$ depends on the form |
| | if the lengths of the sides are $1 : 2 : 3$ , then $c = 8.98$ |

Since the edges of the iceberg become round during melting, we think, that $c = 6$ is a good estimate of this factor.

**Assumption 5** *The form of the iceberg does not change during melting; the factor c in (1) is equal to 6 during the whole towing.*

## 2.2 Mathematical model of the problem

We take the following symbols for the mathematical description of the problem :

$u$ ... velocity of the towing vessel $[km/h]$
$s$ ... position of the towing vessel $[km]$
= distance from Antarctic
$C$ ... costs [ £]
$C'$ ... derivation of $C$ according to $s$
$K$ ... daily rental cost for the largest ship ( see Table 1 in Section 1 )
$V$ ... volume of the iceberg $[m^3]$
$V'$ ... derivation of $V$ according to $s$

$l$ ... cubic root of $V$ $[m]$
$l'$ ... derivation of $l$ according to $s$

Using Assumption 1 we can make a function $mr$ of two variables, which describes the melting rate dependent on the towing speed $u$ and the distance $s$ from the Antarctic.

$$\mathrm{mr}(u,s) = \begin{cases} 0.1\frac{s}{1000}\frac{3+u}{4} & \text{if} \quad 0 \le s \le 1000 \\ 0.2\frac{s-1000}{3000}\frac{3+u}{4} + 0.1\frac{3+u}{4} & \text{if} \quad 1000 \le s \le 4000 \\ 0.3\frac{(3+u)}{4} & \text{if} \quad s > 4000 \end{cases} \tag{2}$$

In the same way we get a function $fc$ of two variables, which describes the fuel costs dependent on the volume $V$ of the iceberg and the velocity $u$ :

$$\mathrm{fc}(V,u) = (\lg V - 1)(1.8 + 0.3u) \tag{3}$$

With this functions we can form two differential equations for the volume of the iceberg and the costs.

Differential equation for the volume of the iceberg :

$$\begin{aligned} dV &= -\mathrm{mr}(u,s)cV^{\frac{2}{3}}dt \\ u &= \tfrac{1}{24}\tfrac{ds}{dt} \\ \Rightarrow dV &= -\mathrm{mr}(u,s)cV^{\frac{2}{3}}\tfrac{1}{24u}ds \\ V' &= -\tfrac{1}{24u}\mathrm{mr}(u,s)cV^{\frac{2}{3}} \end{aligned} \tag{4}$$

If we take l instead of V as variable, then we can form an easier equation :

$$\begin{aligned} dV &= V^{\frac{2}{3}}dl \\ \Rightarrow dl &= -\mathrm{mr}(u,s)c\tfrac{1}{24}ds \\ l' &= -\tfrac{1}{24}\mathrm{mr}(u,s)c \end{aligned} \tag{5}$$

Differential equation for the costs :

$$\begin{aligned} dC &= \mathrm{fc}(V,u)ds + \tfrac{K}{24u}ds \\ C' &= \mathrm{fc}(V,u) + \tfrac{K}{24u} \end{aligned} \tag{6}$$

The initial values for (4) or (5), respectively are given by the volume of the iceberg which should be towed to the Persian Gulf. The initial value for (6) is equal to the costs for travelling from the Persian Gulf to the Antartic.

If we take a function for the relative costs of the produced drinking water ( since you can get only 85 % of the ice volume as water, the volume must be multiplied with the factor 0.85 )

$$h_u(s) = \frac{C_u(s)}{0.85V_u(s)} \tag{7}$$

in wich the subscript $u$ means a dependence on the velocity function $u$, then we have to solve the problem

$$\begin{aligned} h_u(9600) &= min \\ 1 \leq u(s) &\leq 5 \end{aligned} \tag{8}$$

This is a problem of optimal control with the control function $u$.

## 2.3 Approximation of the problem

To solve the control problem numerically, we have to approximate the control function $u$ by a vector $(u_1, u_2, ..., u_n)$, $u_i$ is the constant velocity in the i-th way interval $[s_{i-1}, s_i)$, $s_0 = 0$, $s_n = 9600$.

In this way we get a nonlinear optimization problem, whose solution will be described in the next section.

# 3 Optimization

## 3.1 Transformation of a problem of optimal control into a finite optimization problem

The problem of optimal control, derived in the preceding section can be written more generally as:

$$\begin{aligned} x'(s) &= f(s,x,u) && s \in [0,S] \\ x(0) &= x_0 \\ g_j(s,x,u) &= 0 && s \in [0,S], \quad j = 1,\ldots,m_1 \\ g_j(s,x,u) &\geq 0 && s \in [0,S], \quad j = m_1,\ldots,m \\ h(x_u(S)) &= \min_u \end{aligned} \tag{9}$$

with $x : [0,S] \to \mathbb{R}^k$, $u : [0,S] \to \mathbb{R}^l$, $f : [0,S] \times \mathbb{R}^k \times \mathbb{R}^l \to \mathbb{R}^k$, $g_j : [0,S] \times \mathbb{R}^k \times \mathbb{R}^l \to \mathbb{R}$, $j = 1,\ldots m$, $h : \mathbb{R}^k \to \mathbb{R}$. For ease of notation we assume that $l = 1$.

The solution of this problem of optimal control is carried out numerically. Therefore we divide the interval $[0,S]$ into n, say, not necessarily equidistant subintervals $[s_{i-1}, s_i)$, $i = 1,\ldots,n$ with

$$0 = s_0 < s_1 < s_2 < \ldots < s_{n-1} < s_n = S.$$

The control function $u$ is approximated by a piecewise constant function with value $u_i$ in $[s_{i-1}, s_i)$. By an obvious change of notation, i.e., we do not write down the dependency of the involved functions on the state variable $x$ and on $s$ explicitely anymore, $u = (u_1, u_2, \ldots, u_n) \in \mathbb{R}^n$, we arrive at the following, in general nonlinear, optimization problem:

$$\begin{aligned} h(u_1,u_2,\ldots,u_n) &= \min \\ g_j(u_1,u_2,\ldots,u_n) &= 0, && j = 1,\ldots,m_1 \\ g_j(u_1,u_2,\ldots,u_n) &\geq 0, && j = m_1,\ldots,m. \end{aligned} \tag{10}$$

An effective way of obtaining a solution of (10) is in applying a sequential quadratic programming algorithm which will, in short, be described now. We follow [1].

## 3.2 The sequential quadratic programming approach

An important tool in nonlinear programming is the Lagrange function

$$L(u,\lambda) = h(u) - \sum_{j=1}^{m} \lambda_j g_j(u) \tag{11}$$

with Lagrange parameter $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^m$, which is involved in the well-known necessary optimality conditions, i.e. the Kuhn-Tucker conditions for problem (10)

$$\begin{array}{lcll} \nabla_u L(u,\lambda) & = & 0 & \\ g_j(u) & = & 0, & j = 1,\ldots,m_1 \\ g_j(u) & \geq & 0, & j = m_1+1,\ldots,m \\ \lambda_j & \geq & 0, & j = m_1+1,\ldots,m \\ g_j(u)\lambda_j & = & 0, & j = m_1+1,\ldots,m. \end{array} \tag{12}$$

Here, $\nabla_u$ denotes differentiation with respect to the u-variables. A sequential quadratic programming algorithm proceeds from a quadratic approximation of the Lagrange function (11) and a linearization of the constraints. If $u^k$ denotes the k-th estimate for the optimal solution $u^*$ and $B^k$ a symmetric matrix that approximates the Hessian of the Lagrange function, the resulting programming subproblem can be written in the form

$$\begin{array}{lcll} \frac{1}{2} d^t B^k d + \nabla h(u^k)^t & = & min_d & \\ \nabla g_j(u^k)^t\, d + g_j(u^k) & = & 0, & j = 1,\ldots,m_1 \\ \nabla g_j(u^k)^t\, d + g_j(u^k) & = & 0, & j = m_1,\ldots,m. \end{array} \tag{13}$$

The next iterate is given by

$$u^{k+1} = u^k + \alpha^k d^k \tag{14}$$

where $d^k$ denotes the solution of (13) and $\alpha^k$ a steplength parameter which is obtained by minimizing a line search function.

To improve the numerical performance of the algorithm several adjustments have been made:

- the matrix $B^k$ is updated by a rank-two correction using the BFGS-formula

- active index-set strategy
- introduction of an additional variable to avoid inconsistency of the restrictions
- the line search function is choosen as

$$\phi_r(u,\mu) := h(u) - \sum_{j\in J}(\mu_j g_j(u) - \frac{1}{2}r_j g_j(u)^2) - \frac{1}{2}\sum_{j\in K}\frac{\mu_j^2}{r_j} \tag{15}$$

where the index sets $J$ and $K$ are defined by

$$J := \{1,\ldots,m_1\} \cup \{j | m_1 < j \leq m,\, g_j(u) \leq \mu_j/r_j\}$$

$$K := \{1,\ldots,m\}\backslash J.$$

The definition of the penalty parameters $r_j$, $j = 1,\ldots,m$ can be found in [1]. The variables corresponding to the Lagrange multipliers of the $k-$th quadratic subproblem are denoted by $\mu^k$. The steplength parameter $\alpha^k$ is determined by minimizing the one-dimensional optimization problem

$$\varphi^k(\alpha) = \phi_{r^{k+1}}\left(\begin{pmatrix} u^k \\ \mu^k \end{pmatrix} + \alpha \begin{pmatrix} d^k \\ \lambda^k - \mu^k \end{pmatrix}\right). \tag{16}$$

As mentioned, for further details see [1].

# 4 Results of the optimization

## 4.1 Dependency on the number of intervals

We started the optimization program for different numbers $n$ of way intervals. In every way interval the velocity is constant. The solution of the differential equations for the volume and the absolute costs was carried out by means of the trapezidium rule with a step length of 1 km. The minimal values for the relative costs of the produced drinking water and the corresponding absolute costs and the remaining volume of ice after transportation to the Persian Gulf are given in the following table :

| n | costs [£] | volume [$m^3$] | relative costs [$\frac{£}{m^3}$] |
|---|---|---|---|
| 6 | 198,086 | 2,572,322 | 0.0905963 |
| 12 | 197,726 | 2,575,067 | 0.0903350 |
| 96 | 197,474 | 2,576,615 | 0.0901659 |
| 480 | 198,297 | 2,590,737 | 0.0900478 |

Table 4: Results for different numbers of intervals

The remaining volume of the iceberg increases with more intervals, because the difference between the optimal velocity function and our approximation decreases with a larger number of intervals. The whole costs increases too, since the volume of the iceberg is larger.

The relative costs for producing one cubic meter drinking water differ only in the fourth decimal. The relative costs are decreasing, if we split one way interval into parts, that means that we take a multiple of the number of way intervals.

The corresponding optimal velocity vectors are plotted in Figure 1, Figure 2, Figure 3 and Figure 4. The ship starts with the lowest velocity. Then the velocity is increased until it reaches the upper bound at ca. 3000 km distance from the Antarctic.

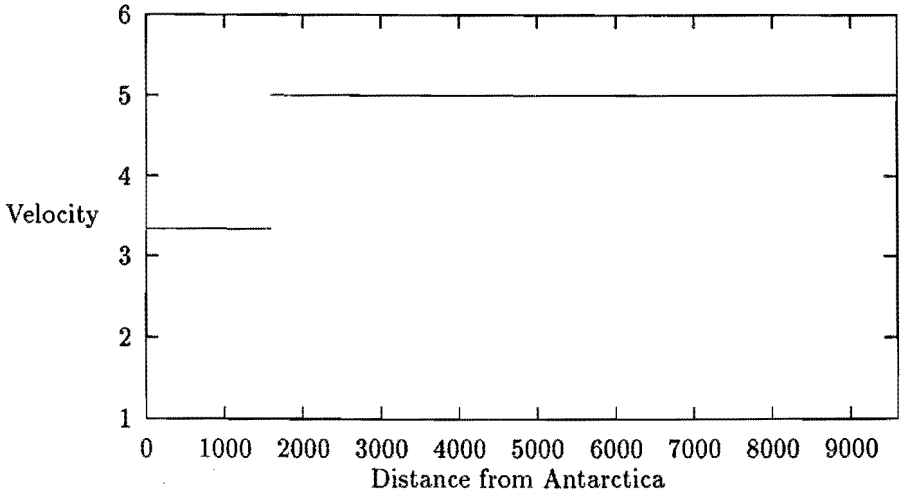



Figure 1: Velocities for 6 way intervals

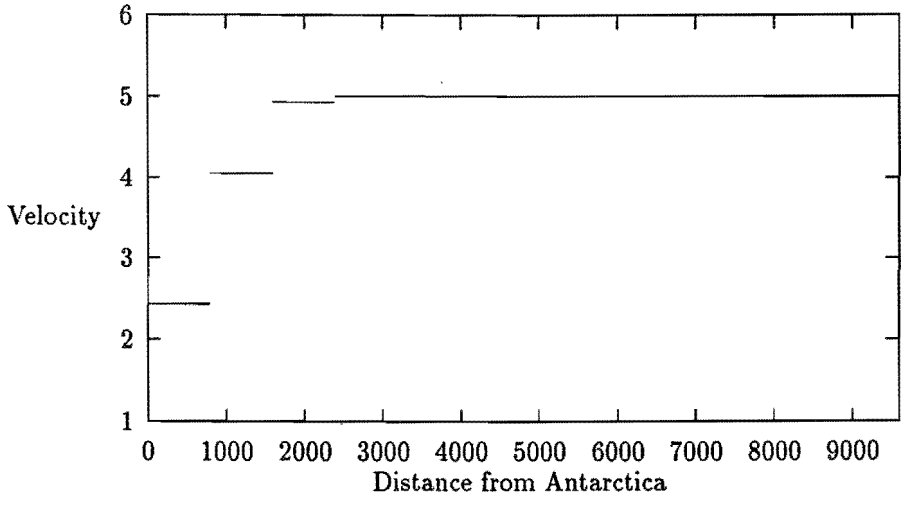



Figure 2: Velocities for 12 way intervals

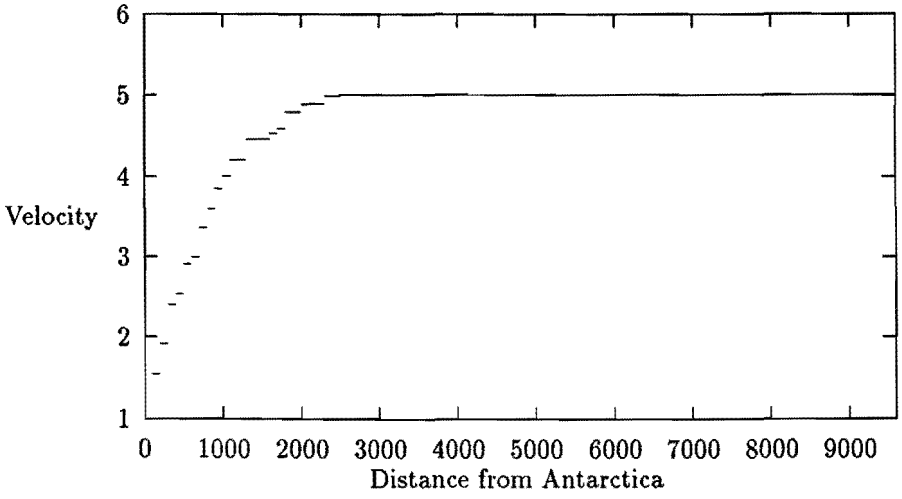



Figure 3: Velocities for 96 way intervals

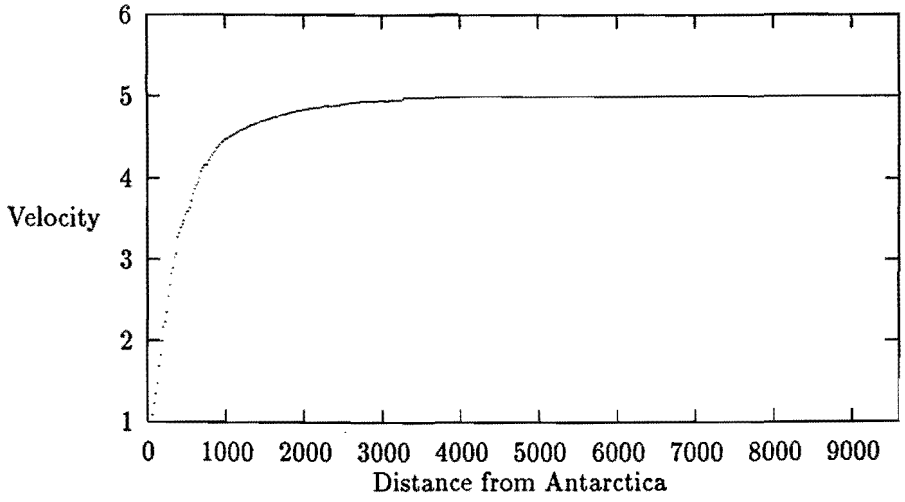



Figure 4: Velocities for 480 way intervals

If the way intervals become smaller, the velocity goes to a continuous function. So the ship must drive with this velocity function to bring the cheapest drinking water to the Persian Gulf.

The volume decreasing and the cost increasing show nearly the same behaviour in all four cases, and therefore we plot here only the curves for 480 way intervals. These curves are plotted in Figure 5 and Figure 6 .

## 4.2 Dependency on the form of the iceberg

In our assumptions for solving the problem we take $c = 6$ as a constant factor between the surface of the iceberg and the cubic root of the volume squared. We suppose that this factor has a value close to 6, but we do not know this value exactly. Therefore we computed for other factors $c$ the optimal solution to see the dependency on this value. For comparison we take a partition in 96 way intervals for all following computations. The results are specified in Table 5 and are plotted in Figure 7.

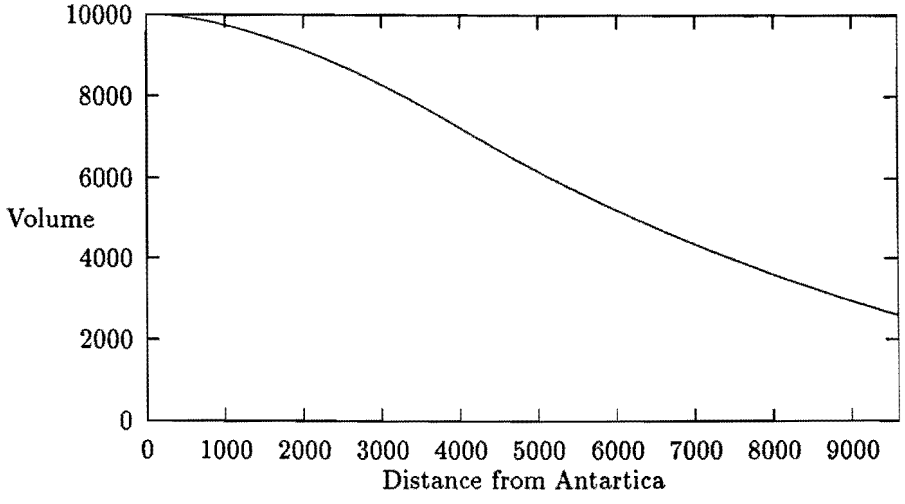



Figure 5: Optimal volume decreasing

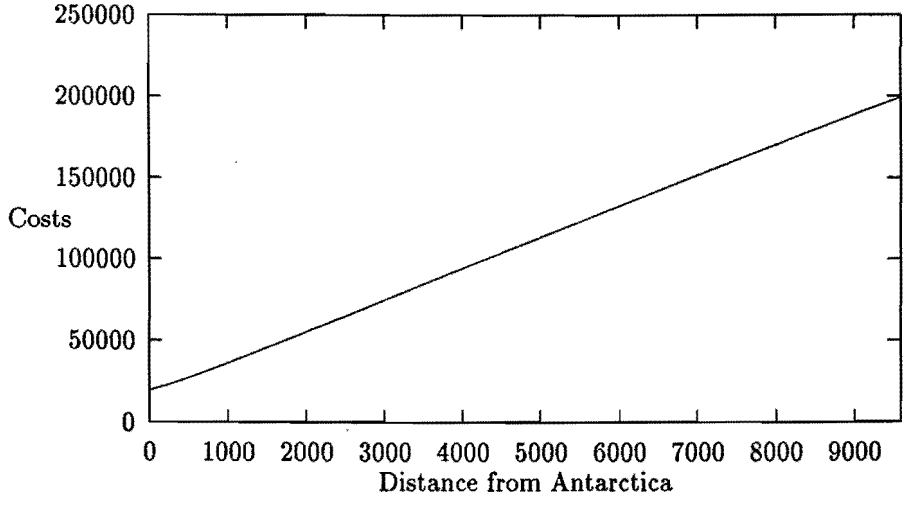



Figure 6: Optimal cost increasing

| c | costs [£] | volume [$m^3$] | relative costs [$\frac{£}{m^3}$] |
|---|---|---|---|
| 4.6 | 196,034 | 3,698,035 | 0.06237 |
| 5.0 | 196,668 | 3,352,875 | 0.06901 |
| 6.0 | 197,474 | 2,576,615 | 0.09017 |
| 7.0 | 197,435 | 1,923,873 | 0.12073 |
| 8.0 | 196,992 | 1,389,549 | 0.16679 |
| 9.0 | 196,055 | 969,966 | 0.31528 |

Table 5: Results for different values of $c$

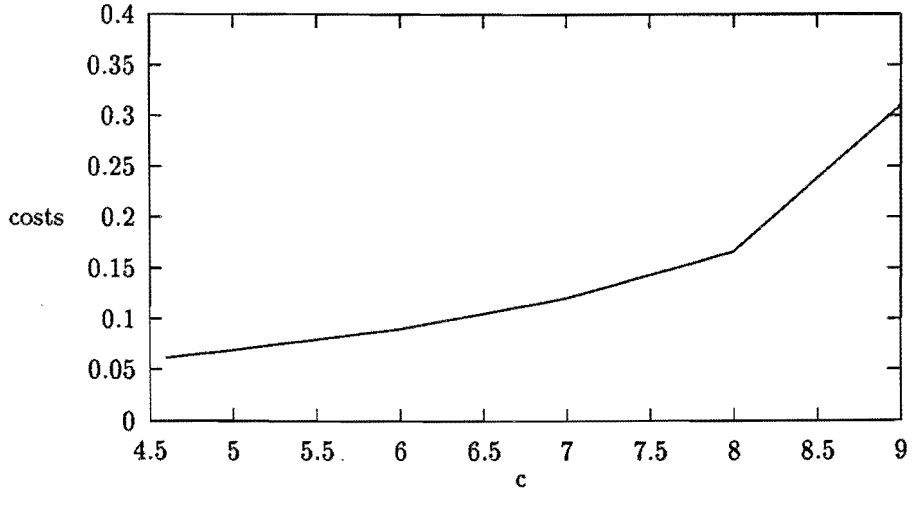



Figure 7: Results for different values of $c$

# 5 Conclusions

If our assumptions are all right, then it is cheaper to tow icebergs from the Antarctic to produce drinking water in the Persian Gulf than desalinating sea water. If we assume, that one human being consumes 50 $l$ water per day, then 120,000 persons can be supplied with drinking water for one year with one remaining iceberg. The needed amount of drinking water for the whole area is larger. So in one year a few ships are necessary to tow icebergs. The exact number can be determined from the exact amount of drinking water consumption.

If the factor $c$ between the surface and the volume is not 6, then the relative costs change. This factor must be experimentally determined. Then it can be decided with the table 5 of the previous section, wether the method is cheaper than the conventional method or not. Approximately we can say, that the critical value is $c_{\text{crit}} = 6.3$. If the real value $c$ is smaller than $c_{\text{crit}}$, then towing icebergs from Antarctic is a good method to produce drinking water.

Since the relative costs are decreasing, if we take larger ships, it must be thought about the possibilities of producing larger ships. If it is possible to build larger ships, then the costs would be lower than our computed values.

We tried to increase the upper bound of the possible velocity from 5 km/h to 7 km/h and got as optimal relative costs 0.0866 $£/m^3$. Therefore another possibility to decrease the costs is the construction of faster towing vessels.

The costs also decrease, if not every ship tows one iceberg, but a few ships tow one big iceberg together, because the relation between surface and volume is smaller for a larger iceberg, and so a smaller part of the iceberg melts. Therefore this is another possibility to decrease the costs. But there are a few questions :

- How many ships can together tow an iceberg ?
- What is the limit for the size of the icebergs to may be melted in the Persian Gulf ?

The answers to these questions give a lower bound on the costs.

The distance between Australia and the Antarctic is very small compared with the distance to the Persian Gulf. Australia has a large desert area. So we think this method is not only applicable for the Persian Gulf. Producing drinking water for Australia with this method is very cheap.

# TRAFFIC LIGHTS

# Traffic lights

## Contents

# 1 Introduction

On a road one of the two lanes is blocked over a distance of 500 meters due to maintenance. We are supposed to regulate the traffic. To do that there are traffic lights placed on both sides of the road block. There is a speed limit of 20 $(km/h)$ at the free lane. From side 1 arrive approximately 800 cars per hour, from the other side (side 2) 300 cars.

In this paper we look at two models to regulate the traffic. At first we consider a simple model, where we deduce two linear equations with the unknown "green times" for the two traffic lights. We solve these equations and compare the results with the numerical simulations of an extended model. This extended model can't be solved analytically, due to randomness of different variables, i.e. the flow of the approach, the reaction time, and the length of a car.

The second model is based on queueing theory. The expected minimum waiting time is minimized for two different strategies, i.e. fixed green and red times for both traffic lights, and green times depending on the number of cars that are in the queue.

# 2 The first model, a direct approach

We make the following assumptions:

- the flow of the cars is uniform;
- a = 2 $(m/s^2)$, when cars leave the queue;
- the reaction time for every car in the queue is the same;
- the average length of a car is 5 $(m)$;
- the average distance between two cars in a queue is 1 $(m)$;

These assumptions will be weakened at the end of this paragraph, when we make some numerical simulations.

We distinguish the four following periods of the traffic lights. Four following periods are called a cycle and are repeated time after time.

| time | traffic light on side 1 | traffic light on side 2 |
|---|---|---|
| $t_1$ | green (G) | red (R) |
| $t_0$ | red | red |
| $t_2$ | red | green |
| $t_0$ | red | red |

The time $t_0$ is the time that it takes a car to drive the distance between the traffic lights. This distance is 500 m, and the maximum speed is 20 km/h. So a car travels the distance in 90 s. Adding a safety margin of 3 seconds gives $t_0 = 93$ s.

Now that we have determined $t_0$, the problem is to determine $t_1$ and $t_2$. To do this we first write down the number of cars that arrive and depart in every period of a cycle. (Note that 800 cars/h = 2/9 cars/s and 300 cars/h = 1/12 cars/s.) This gives the following table, in which N(t) denotes the number of cars that pass the traffic light in time period t.

| arrival | departure | t | | | departure | arrival |
|---|---|---|---|---|---|---|
| $\frac{2t_1}{9}$ | N($t_1$) | G | $t_1$ | R | 0 | $\frac{t_1}{12}$ |
| $20\frac{2}{3}$ | 0 | R | 93 | R | 0 | $7\frac{3}{4}$ |
| $\frac{2t_2}{9}$ | 0 | R | $t_2$ | G | N($t_2$) | $\frac{t_2}{12}$ |
| $20\frac{2}{3}$ | 0 | R | 93 | R | 0 | $7\frac{3}{4}$ |

In one green period we will try to solve the queue that arises in one cycle. Otherwise queues will start to grow. So we have to solve the following inequalities, in which $t_1$ and $t_2$ and N(t) are unknowns:

$$41\frac{1}{3} + \frac{2t_1}{9} + \frac{2t_2}{9} \leq N(t_1) \tag{1}$$

$$15\frac{1}{2} + \frac{t_2}{12} + \frac{t_2}{12} \leq N(t_2) \tag{2}$$

The next step is to determine N(t). Note that N(t) can be measured by just counting the number of cars that pass a traffic light which suffices our requirements. Due to the fact that we don't have such a traffic situation, we make some assumptions to "approximate" the value of N(t). We assume that the accelaration by departure from the queue is 2 $m/s^2$, and we know that the maximum velocity is 20 km/h = $5\frac{5}{9}$ m/s. So:

$$s(t) = \begin{cases} t^2 & \text{if } 0 < t < 2\frac{7}{9} \\ 7\frac{58}{81} + 5\frac{5}{9}(t - 2\frac{7}{9}) & \text{if } t > 2\frac{7}{9} \end{cases} \tag{3}$$

We assume that every car has the same reactiontime $\epsilon$, so the distance travelled by car i equals $S_i$(t) = s(t-i$\epsilon$). The distance of the i-th car in the queue to the traffic light is $d_i$ = (i-1)l, where l=6m , that is the sum of the average length of a car and the average distance between two cars. Equalling $d_i$ and $S_i(t)$, and rewriting this for a given t, gives the number of cars that pass the traffic light during time t. So we have determined N(t).

$$N(t) = \frac{13.72 + 5.56(t - 2.78)}{6 + 5.56\epsilon} \tag{4}$$

Substitution of N(t) in (2) with a reactiontime of 0.5 s gives two inequalities with two unknowns. The solution of these inequalities are the minimal values which prevent the queues from growing:

$$t_1 \geq 127.0 \tag{5}$$

$$t_2 \geq 47.8 \tag{6}$$

## 2.1 Numerical simulation

Now that we know the minimal values for the deterministic model, we look at a more realistic model, with a Poisson arrival process. We also use some stochastic parameters concerning the arrival of cars. Again we note that this couldn't be necessary if we could count the number of cars that pass a traffic light which suffices our requirements.

In our program we computed the minimal green times $t_1, t_2$ , which prevent the queues from growing. The simulation gave the following optimal values for the green-times:

$$t_1 = 139s \tag{7}$$

$$t_2 = 52s \tag{8}$$

For these values we calculate the average waiting time $EW$:

$$EW = \frac{\lambda_1}{\lambda_1 + \lambda_2} EW_1 + \frac{\lambda_2}{\lambda 1 + \lambda 2} EW_2 \tag{9}$$

with $EW$ the average waiting time, and $EW_i$ the average waiting time for queue i.

First we determine the waiting time for all the cars in queue 1, in which $\epsilon$ denotes the reaction time:

| car number in queue | waiting time |
|---|---|
| 1 | $t_2 + 2t_{rr} + \epsilon$ |
| 2 | $t_2 + 2t_{rr} + \epsilon - 1/\lambda_1$ |
| 3 | $t_2 + 2t_{rr} + \epsilon + 2(-1/\lambda_1 + \epsilon)$ |
| . | . |
| . | . |
| n | $t_2 + 2t_{rr} + \epsilon + (n-1)(-1/\lambda_1 + \epsilon) = \epsilon,$ |

All the waiting times added gives

$$EW_1 = \frac{t_2 + 2t_{rr} + \epsilon + \epsilon}{2} \tag{10}$$

Similarly for queue 2:

$$EW_2 = \frac{t_1 + 2t_{rr} + 2\epsilon}{2} \tag{11}$$

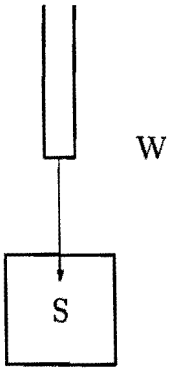
Substituting the times found by the numerical simulation program gives an average waiting time of 131.4 s.

# 3 The second approach, using queueing theory

## 3.1 Introduction

In our second approach we make use of some queueing theory. First this is explained, then we make some assumptions. After this is done a formula is given, which gives an approximation for the average waiting time of a queueing system with service interuptions. This formula is valid for one queue and is derived in [3]. At first sight, one can obtain a formula for the average waiting time for two queues. But, as we see, this approach won't work. In the third paragraph we use another kind of traffic light. This traffic light hasn't got fixed green and red times, but the red and green times depend on the lengths of the queues.

We use a queueing system with service interuptions. In a queueing system there is a server, $S$, which serves customers who wait in an infinitely large waiting room, **W** (see figure below). The flow of the approach of the customers is determined by a Poisson ($\lambda$) process. After the server served the customer, which takes a time, S, the customer leaves the waiting room $W$. The customer waits time W in the waiting room. But there is a restriction. There are service interuptions during which the server is closed over a fixed time $\beta$. The time the server serves the customers is $\alpha$. So it is obvious, that if there are no service interruptions then $\beta = 0$.




We define the following entities:

- $E(S)$ = average service time,
- $E(W)$ = average waiting time,

- $\rho$ = traffic intensity,

and we will use Little's formula:

$$\rho = \lambda E(S). \tag{12}$$

We will apply this theory to our problem. But first we make the following assumptions:

- all cars pass the traffic light with 20 km/h;
- the service time S, defined by the time it takes for a car plus the distance between the car and the subsequent car to pass the green light, has for both queues the same distribution with first moment $E(S) = 1.58$ and second moment $E(S^2) = 0.54$.

With the first assumption the second is equivalent to the assumption that the distance between the fronts of two subsequent driving cars has a mean which is approximately equal to 9 m. The second assumption is not valid for the first two or three cars in the queue because they first have to accellerate, the influence of this on the final result will not be large.The service time S is approximately equal to the sevice time in chapter two.

We get the following result out of this theory immediately. Because we want the number of cars which arrive during one cycle to be less than the number of cars which depart in one green time ( otherwise it is certain that the queues will grow ) for both queues, we get the following inequalities:

$$\lambda_i(\alpha_i + \beta_i) < \frac{\alpha}{E(S)} \qquad i = 1, 2 \tag{13}$$

We determine minimal green times out of these inequalities. They are:

$$t_1 \geq 127s. \tag{14}$$

$$t_2 \geq 48s. \tag{15}$$

## 3.2 Using the Fischer approximation

According to [3], the following approximation is valid when $\rho$ is close to 1:

$$E(W) = \frac{-V}{2M}, \tag{16}$$

where

$$M = \rho - \frac{\alpha}{\alpha + \beta}, \tag{17}$$

$$V = \lambda E(S^2). \tag{18}$$

Further we define:

$W_1$ = average waiting time queue 1

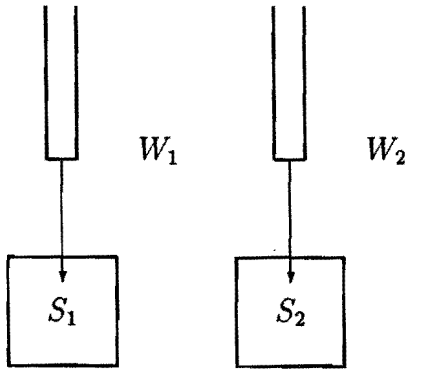
$W_2$ = average waiting time queue 2.

We want to minimize the following expression over $t_1$ and $t_2$:

$$W = \frac{1}{\lambda_1 + \lambda_2}(\lambda_1 W_1 + \lambda_2 W_2) \tag{19}$$

i.e. the weighted mean of the average waiting times for both queues.

Now we have the following situation:




$$\alpha_1 = t_1$$

$$\alpha_2 = t_2$$

$$\beta_1 = t_2 + 2t_0$$

$$\beta_2 = t_1 + 2t_0$$

The discription on the last page is not the actual situation, because here we have two servers $S$, and in practice we only have one server $S$.

In order to minimize W we derive the partial derivatives of W. After equating the partial derivatives to zero, we get this remarkable result:

$$t_1 + t_2 + 2t_0 = 0 \tag{20}$$

Since this is not possible for three positive numbers, it appears that we have made a mistake. This is probably due to the fact that we modelled a different situation. We looked for a better way to model the traffic light problem using [3], but all these models didn't work out.

## 3.3 Using a cyclic, single-server queueing theory

The problem of regulating traffic lights can be solved with queueing theory. In [1] and [2] the mean waiting times for cyclic single-server systems are calculated. For our problem we use this theory because we also have one server, namely the traffic light, and because the system is cyclic. After serving one lane the other is served and then the first lane is served again, and so forth.

### 3.3.1 General Model

We will shortly describe the general model presented in [1] and [2].Consider $N$ queues $Q_1, ..., Q_N$; all queues have infinite capacity. At all queues cars arrive according to independent Poisson processes with arrival intensities $\lambda_1, ..., \lambda_N$. The total arrival rate is given by:

$$\lambda = \sum_{i=1}^{N} \lambda_i \tag{21}$$

The queues are attended by a single-server $S$ who visits the queues in a fixed cyclic order: $Q_1, Q_2, ..., Q_N, Q_1, Q_2, ....$ The switch-over times of the server between the $i$th and $(i+1)$th queue are independent, identically distributed stochastic variables with first moment $m_i$, and second moment $m_i^{(2)}$. The first moment of the total switch-over time during a cycle of the server, $m$, is given by:

$$m = \sum_{i=1}^{N} m_i \tag{22}$$

The second moment of m is given by $m^{(2)}$.

The service times of cars in lane $i$ are independent, identically distributed stochastic variables. Their distribution $S_i(.)$ has first moment $s_i$ and second moment $s_i^{(2)}$. It is assumed that the interarrival process, the service process and the switch-over process are mutually independent. The offered traffic at $Q_i$, $\rho_i$, is defined as:

$$\rho_i = \lambda_i s_i, \qquad i = 1, ..., N. \tag{23}$$

The total offered traffic, $\rho$, is defined as:

$$\rho = \sum_{i=1}^{N} \rho_i, \tag{24}$$

$\rho$ is the average amount of service time that can be offered in a time unit. Per time unit maximal one time unit of work can be done. So a necessary condition on $\rho$ is: $\rho < 1$.

If we define:

$EM_i^{(1)}$ : the mean amount of work in $Q_i$ at a departure epoch of the server $S$ from $Q_i$,

then it can be shown that the following relation for the average waiting time in the queues, $EW_i$, holds:

$$\sum_{i=1}^{N} \rho_i EW_i = \rho \frac{\sum_{i=1}^{N} \lambda_i s_i^{(2)}}{2(1-\rho)} + \rho \frac{m^{(2)}}{2m} + \frac{s}{2(1-\rho)} \left[ \rho^2 - \sum_{i=1}^{N} \rho_i^2 \right] + \sum_{i=1}^{N} EM_i^{(1)} \tag{25}$$

The last three terms reflect the influence of the presence of switchover times. The term $\rho m^{(2)}/2m$ represents the mean amount of work that arrived at all queues during the switching intervals after the last visit of the server $S$ to those queues. The next term reflects the interaction between queues, it represents the mean amount of work that arrived at queues, after the last visit of $S$, during the subsequent service periods of other queues. The last term represents the mean total amount of work left behind by $S$ at the various queues in one cycle. This is the *only* term that cannot be determined without specifying the service strategies at the queues. In [2] eight possible service strategies are considered. We will not discuss them here. In the next paragraph we describe only the strategy that can be used in our problem.

### 3.3.2 Application to the traffic light problem

In our problem we only have two queues with infinite capacity, and the server is the traffic light. The arrivals of cars are assumed to be distributed Poisson with mean $\lambda_1 = 1/4.5$ for lane 1, and $\lambda_2 = 1/12$ for lane 2. So,

$$\lambda = \frac{1}{4.5} + \frac{1}{12} = \frac{11}{36}$$

The switch-over time in our case is equal to the time that both traffic lights are red. This is a fixed time, so

$$\begin{aligned} m_1 = m_2 = 93 \quad &\rightarrow \quad m = 186\ (sec) \\ m_1^{(2)} = m_2^{(2)} = 93^2 \quad &\rightarrow \quad m^{(2)} = 93^2 \end{aligned}$$

For safety reasons a small period of time can be added to $m$.

The service time of a car is the time it takes a car to pass the traffic light. As in section 3 we take for the average service time and the second moment of both queues:

$$\begin{aligned} s = s_1 = s_2 = \ & 1.58 \\ s^{(2)} = s_1^{(2)} = s_2^{(2)} = \ & 0.54 \end{aligned}$$

The offered traffic at lane 1 is $\rho_1 = \lambda_1 s_1 = 1/4.5 \times 1.58 = 0.35$ and $\rho_2 = 1/12 \times 1.58 = 0.13$. The condition $\rho < 1$ then holds, because

$$\rho = \rho_1 + \rho_2 = 0.48 < 1.$$

We think there are two possible service strategies for a traffic light. The first strategy, which is also used in section 2.1, is green periods of fixed time. However, for this strategy it has not been possible to determine the corresponding $EM_i^{(1)}$'s yet.

The second strategy is called gated strategy. At the moment the traffic light turns to green it checks how many cars are in the queue. These cars may pass the traffic light and then it turns red. The mean amount of work in queue $Q_i$ for this strategy is:

$$EM_i^{(1)} = \rho_i \frac{m}{1-\rho}.$$

and equation (25) becomes

$$\begin{aligned}\sum_{i=1}^{2} \rho_i EW_i &= \rho \frac{\sum_{i=1}^{2} \lambda_i s_i^{(2)}}{2(1-\rho)} + \rho \frac{m^{(2)}}{2m} + \frac{m}{2(1-\rho)} \left[\rho^2 - \sum_{i=1}^{2} \rho_i^2\right] + \sum_{i=1}^{2} \rho_i \frac{m}{1-\rho} \\ &= \rho \frac{\sum_{i=1}^{2} \lambda_i s_i^{(2)}}{2(1-\rho)} + \rho \frac{m^{(2)}}{2m} + \frac{m}{2(1-\rho)} \left[\rho^2 + \sum_{i=1}^{2} \rho_i^2\right] \qquad (26)\end{aligned}$$

Therefore:

$$0.35EW_1 + 0.13EW_2 = 0.076 + 22.32 + 66.137 = 88.53 \qquad (27)$$

From this we can obtain the average waiting time, $EW$, in both queues by dividing by $\rho$:

$$\begin{aligned}EW &= \frac{\rho_1 EW_1 + \rho_2 EW_2}{\rho} = \frac{\lambda_1 s EW_1 + \lambda_2 s EW_2}{\lambda_1 s + \lambda_2} \\ &= \frac{\lambda_1}{\lambda_1 + \lambda_2} EW_1 + \frac{\lambda_2}{\lambda_1 + \lambda_2} EW_2 \qquad (28)\end{aligned}$$

Thus the average waiting time for a car in an arbitrary queue is approximately 3 minutes and 4 seconds or:

$$EW = \frac{88.53}{0.48} = 184.4 \tag{29}$$

Note that this average time is longer than for the fixed green and red times. This is probably due to more switch over times when using the gated strategy. In practice it must be possible to use this service strategy. For instance, this can be done by placing sensors in the road. At a certain distance before the traffic light, a sensor starts to count how many cars get queued at the moment the light turns green. After one cycle the sensor knows exactly how many cars are in the queue. A sensor at the traffic light then counts the number of passing cars. At the moment all the queued cars passed, the sensor sends to the traffic light, which turns red immediately. The only problem is that it is not easy to define at which distance the first sensor should be placed. A good position can be obtained by trial and error.

# 4 Conclusions and Recommendations

We recommend to use the smart traffic lights, as described in the last section, with the gated strategy. The average waiting time is then approximately 3 minutes. The reason for this is obvious. Because this kind of traffic lights is more flexible, the light is longer green if there is a larger queue. But if it isn't possible to use these traffic lights due to practical reasons then the best alternative is to use normal traffic lights, with the following green and red times: $t_1 = 139$ s., $t_2 = 52$ s. and $t_0 = 93$ s.

# References

[1] O.J. Boxma and W.P. Groenendijk
Pseudo-conservation laws in cyclic-service systems
*Journal of Applied Probability*, vol. 24, pp. 949-964, 1987

[2] O.J. Boxma
Workloads and Waiting Times in Single-Server Systems with Multiple Customer Classes
*Centre for Mathematics and Computer Science*, Report BS-R8906, March, 1989

[3] M.J. Fischer
An approximation to queueing systems with interuptions
*Management Science*, vol. 24, no. 3, pp. 338-344, 1971