TA ed.

FORSCHUNG - AUSBILDUNG - WEITERBILDUNG

BERICHT NR. 28

PATTERN RECOGNITION USING MEASURE SPACE METRICS

HELMUT NEUNZERT, BRIAN WETTON

UNIVERSITÄT KAISERSLAUTERN
FACHBEREICH MATHEMATIK
ERWIN-SCHRÖDINGER-STRASSE
D - 6750 KAISERSLAUTERN

NOVEMBER 1987

ABSTRACT

Patterns are considered as normalized measures and distances between them are defined as distances of the corresponding measures using metrics in measure spaces. This idea can be applied for pattern recognition if "smeared" patterns have to be compared with given ideal patterns. Different metrics are sensitive to different characteristics of the patterns - this is demonstrated in discussing examples. Particular attention is paid to a problem of Qualilty Control for an artificial fabric, where the distance to uniformity is defined and evaluated; the results are now used in industry.

I INTRODUCTION

We would like to introduce a few new thoughts on the problem of Pattern Recognition and to explain them using examples. We were led to a consideration of this topic by practical problems in Quality Control of irregularities in the thickness of an artificial fabric, and in automatic recognition of paper currency. Both of these problems were posed to us by Industry. The mathematical concepts that we are using have not, to our knowledge, been applied to Pattern Recognition and seem to be well suited to at least some special types of problems. The following two questions arose:

1. We are given a finite set of ideal patterns (perfect bank notes, for example) and a smeared pattern (an old, well-used note). We want to identify the old bank note as one of the possible kinds. This is a 'classical' problem of Pattern Recognition. Some standard procedures for solving this problem can be found in [1].

2. We are given a finite set of smeared or noise-contaminated patterns (sections of fabric) and an ideal pattern (a fabric with uniform thickness). The question is which pattern is closest to the ideal pattern (i.e. has the smallest irregularity)?

Both questions can clearly be answered if we had a sensible concept of distance between patterns: In the first case, we identify the smeared pattern as the ideal one that it is "nearest" to; in the second case, a noisy pattern is considered to be better the "closer" it is to the ideal.

The fundamental idea of this paper is that we interpret patterns as measures in the mathematical sense and then define the distance between patterns using a metric on the space of measures. There are many different metrics on the space of (normalised or probability) measures and they are all sensitive to different characteristics of the pattern. The trick in the mathematical modelling is to find the appropriate metric for every application. In this report, we will present a few different metrics and discuss what practical situations they are best suited to. We pay particular attention to the problem of fabric uniformity. A suitable metric for this problem and an algorithm for computations in the case of discrete measurements is presented.

II BASIC IDEAS AND NOTATION

We consider the set P of all Borel probability measures on Ω, a bounded set in $\mathbb{R}^n$. An element μ of P represents a possible pattern. For example, if Ω were the interval [0,1], $\mu \in P$ could be the mass distribution on 1 meter of fabric (we assume a one-dimensional dependence). In practice, μ is determined by the amount of absorbtion of a laser beam as it scans a one-dimensional track in the fabric. Our general problem is to examine different metrics ρ on Ω, so that we can consider the distance $\rho(\mu,\nu)$ between two patterns μ and ν.

In the case of our quality control problem, we want to find a function $i(\mu)$ representing the irregularity of a pattern

μ in P. Since the irregularity can be thought of the distance from the ideal pattern μ^0 which is the uniform distribution in Ω, we define $i(\mu)$ as $\rho(\mu,\mu^0)$. When $\Omega=[0,1]$ we set μ^0 is equal to the Lebesque measure.

In practice, however, we rarely have continuous data (the laser scanner measures the absorbtion at discrete points). We will consider the following discrete case in applications: $P'(n)$ is the set of probability distributions on the set $\{1,2, \ldots n\}$, where the points are thought to correspond to points $\{x_1,x_2, \ldots x_n\}$ of Ω. $\mu \in P'$ is then given by an element $(\mu_1, \mu_2, \ldots, \mu_n)$ of $[0,1]^n$ such that $\Sigma\mu_i=1$ and may also be interpreted as a measure $\bar{\mu}$ on Ω as follows:

$$\bar{\mu} = \sum_{i=1}^{n} \mu_i \delta_{x_i}.$$

Here δ_x is the usual δ measure at x. In regular scanning of $[0,1]$, $x_i = i/n$. In our quality control example, μ_i would then be the scaled (so that $\Sigma\mu_i=1$) absorbtion at these points.

For the case $\Omega=[0,1]$, $x_i = i/n$, and $\mu \in P'$, we define the discrete irregularity function $i'(\mu)$ as $\rho(\bar{\mu},\bar{\mu}^0)$ with μ^0 given by $\mu^0{}_i= 1/n$ for $i=1,2, \ldots n$.

Let us now consider a first example of a metric on P, derived from the total variation norm $\|\cdot\|_V$ in a standard way: $\rho_V(\mu,\nu):= \|\mu-\nu\|_V$. The total variation norm is defined on finite signed measures on Ω as follows:

$$\|\mu\|_V = \sup \sum_{i=1}^{\infty} |\mu(E_i)|$$

where the sup is taken over all countable Borel partitions

$\{E_i\}$ of Ω. This metric defines the irregularity function i_V'. It is clear that $i_V'(\mu)$ is equal to the $L_1(\mathbb{R}^n)$-norm of $(\mu_1-1/n, \mu_2-1/n, \ldots \mu_n-1/n)$. Figure 1 shows two test examples that will be used during the course of this work. Table 1 shows the performance of i_V' on these examples. We see now why this is a poor measure of irregularity: The two patterns have the same irregularity but pattern A is considerably worse in the sense that it contains a rather large hole in it. The problem is that the order of μ_i's does not play any role, that the method does not realize that a bunch of small holes in the fabric are not as bad as one large one. The same problem occurs if we try to compute the distance between μ and μ^0 using any of the other usual L_p-distances on $\mathbb{R}^n$.

We are not restricted to metric functions on P. Since we are interested in "disorder", it may be tempting to use the concept of entropy, which measures order and disorder in statistical mechanics:

$$E(\mu|\mu^0) := \sum_{i=1}^{n} \mu_i \cdot \ln(\mu_i/\mu^0{}_i)$$
$$= \ln(n) + \sum_{i=1}^{n} \mu_i \cdot \ln(\mu_i) \qquad (2.1)$$

Here, $E(\mu|\mu^0)$ is the entropy of μ relative to μ^0 and in the second line above we have taken μ^0 to be the uniform distribution defined above. Since $E(\mu|\mu^0)$ is zero when $\mu=\mu^0$ and attains its maximum ln(n) when $\mu_i=\delta_{ij}$ for some j, we can consider $E(\mu|\mu^0)$ as a measure of how different μ is from μ^0 and so how "bad" it is. However, one need only look at the formula for Entropy (2.1) to see that it is also not influenced by the order of the measurements. Patterns A and B would be considered equally good by this method, which we do not want.

In the next sections, we look at other metrics on P and the irregularity functions that they generate.

General information on probablility measures can be found in [2] and [3].

III BOUNDED LIPSCHITZ DISTANCE

Let D be the set of real valued functions f on [0,1] such that $f(x)\in[0,1]$ and $|f(x)-f(y)|\leq |x-y|$ for all $x,y\in\Omega$. The Bounded Lipschitz Distance between two elements μ and ν of P can be written as:

$$\rho_b(\mu,\nu) = \sup_{f\in D} \left|\int f\cdot d\mu - \int f\cdot d\nu\right|. \qquad (3.1)$$

The integration is taken over Ω. Since the addition of a constant to f does not change the value of the expression within the sup in (3.1), the condition $f(x)\in[0,1]$ can be dropped when the diameter of Ω is less than or equal to one (when $\Omega=[0,1]$ for instance). In this case, the discrete irregularity i_b' associated with this distance can now be defined and then reduced to an easier form:

$$\begin{aligned} i_b'(\mu) &:= \rho_b(\bar{\mu},\bar{\mu}^0) \\ &= \max_{f\in D} \left|\sum_{i=1}^{n} f_i\cdot(\mu_i-1/n)\right| \qquad (3.2) \end{aligned}$$

where $D' := \left\{f\in\mathbb{R}^n:\ |f_i-f_{i+1}|\leq 1/n\right\}$. We can set $f_1\equiv 0$ without loss of generality since the addition of a constant to f does not change the value of the expression in the maximum above. See Table 1 for an application of this technique to the examples in Figure 1. We see the properties that we want, that pattern B is more irregular than A.

The computation of the irregularity i'_b can be achieved using the simplex algorithm since (3.2) defines a linear optimisation problem with inequality contraints.

When $\Omega=[0,1]$, the Bounded Lipschitz Distance is identical to the following Kantorowitsch and Rubinstein Metric:

$$\rho_{kr}(\mu,\nu) := \inf_{\psi\in\Psi} \int |x-y| \, d\psi$$

where Ψ is the set of all positive borel measures on Ω^2 such that $\psi(E\cdot\Omega)-\psi(\Omega\cdot E) = \nu(E)-\mu(E)$. Since the metrics are the same, the discrete irregularity functions they generate are identical, although the different representation gives rise to a different numerical procedure. A detailed account of the Kantorowitsch-Rubinstein Metric is given in [4].

The Bounded Lipschitz Distance has another interesting mathematical property which has practical implications: it defines the weak convergence in P, i.e. $\lim_{n\to\infty} \rho_b(\mu_n,\mu)=0$ iff $\mu_n \xrightarrow{W} \mu$. Physically, this means that the ρ_b-distance between a pattern (say the letter A) and the same pattern slightly smeared or shifted (the letter A read by an optical reader slightly out of alignment) is small. One can calculate the following directly from (3.1): $\lim_{x\to 0} \rho_b(\mu,\mu_x) = 0$, where μ_x is the measure μ shifted by x. The application of this idea is being investigated by the authors.

IV DISCREPANCY

In the case of Bounded Lipschitz Distance, discrete irregularity computions involved a linear optimisation problem with n-1 state variables and 2n-2 inequality constraints. In our application, we wish to consider n=4096. Clearly, the calculation time for this method is prohibitive. We now consider Discrepancy, which does not suffer from this drawback.

The Discrepancy metric for the case $\Omega=[0,1]$ is defined as follows:

$$\rho_*(\mu,\nu) := \sup_{0\leq\alpha\leq\beta\leq 1} |\mu([\alpha,\beta])-\nu([\alpha,\beta])|$$

The idea of Discrepancy was first introduced in a number theoretical setting by Hermann Weyl (see [5]). It is still the object of considerable interest in such diverse fields as Numerical Integration and Diophantine Approximation [6]. The discrete irregularity corresponding to this metric is

$$i_*^!(\mu) := \max_{1\leq\alpha\leq\beta\leq n} \left| \sum_{i=\alpha}^{\beta} (\mu_i - 1/n) \right| . \qquad (4.1)$$

The results of this method are shown in Table 1. One can see directly from the formula (4.1) that $i_*^!(\mu)$ measures the largest "hole" in μ, where we consider positive and negative deviations around μ^0 to be holes.

The question now is how do we calculate, $i_*^!$. A norm on $\mathbb{R}^n$ is closely related to the calculation of $i_*^!$. The norm is defined as

$$\| f \| := \max_{1\leq\alpha\leq\beta\leq n} \left| \sum_{i=\alpha}^{\beta} f_i \right| . \qquad (4.2)$$

We must only find a way to evaluate this norm for any f since $i_*^!(\mu)=\|\mu-\mu^0\|$. We turn first to a direct evaluation of (4.2). We define:

$$F_0 := 0$$
$$F_j := \sum_{i=1}^{j} f_i \ .$$

Then

$$\|f\| = \max_{0\leq\alpha<\beta\leq n} |F_\beta - F_\alpha| \ .$$

The calculation of F and $\|f\|$ as above is of order n^2.

In order to describe a better algorithm, we will need a few notations and ideas.

Let $n>2$. $f\in\mathbb{R}^n$ is called alternating when $f_i \cdot f_{i+1} \leq 0$ for $i=1,2,$ ... n. We write $f\in\mathbb{R}_a^n$.

If, moreover, for some $\alpha\in\{1,2, \ldots n-2\}$ we have $|f_\alpha|\geq|f_{\alpha+1}|$ and $|f_{\alpha+2}|\geq|f_{\alpha+1}|$ then we call $f\in\mathbb{R}_a^n$ contractable.

A contraction $\check{f}$ of a contractable f is defined as follows:

$$\check{f}_i := \begin{cases} f_i & \text{when } i<\alpha \\ f_\alpha+f_{\alpha+1}+f_{\alpha+2} & \text{when } i=\alpha \\ f_{i+2} & \text{when } i>\alpha \end{cases}$$

where $\alpha(f) := \min \{i: |f_i|\geq|f_{i+1}| \text{ and } |f_{i+2}|\geq|f_{i+1}|\}$

It is clear that $\check{f}\in\mathbb{R}_a^{n-2}$.

For $f\in\mathbb{R}^n$ we now denote by m(f) the number of alternations of f and by $I(f)_j$ the index of the j-th alternation:

$$m(f) := |\{i: f_i \cdot f_{i+1} \leq 0\}|$$

$$I(f)_j := \min \left\{ i: f_i \cdot f_{i+1} \leq 0 \text{ and } i > I(f)_{j-1} \right\}$$

We write $I(f)_0=1$ and $I(f)_{m(f)+1}=n$ for convenience.

We now group the alternations together and define $\hat{f} \in \mathbb{R}^{m(f)}$ by

$$\hat{f}_j := \sum_{i=I(f)_{j-1}}^{I(f)_j} f_i, \quad i=1,2, \ldots m(f)+1$$

It is clear that $\|f\|=\|\hat{f}\|$ and $\hat{f} \in \mathbb{R}_a^{m(f)}$ for all $f \in \mathbb{R}^n$.

Lemma 1: If $f \in \mathbb{R}_a^n$ and is not contractable, then

$$\|f\| = \max_{1 \leq i \leq n} |f_i|$$

Proof: We assume that $\|f\| > \max |f_i|$. It follows then that

$$\|f\| = \left| \sum_{i=l}^{m} f_i \right|$$

for certain l and m with $l-m \neq 0$. We assume without loss of generality that

$$\sum_{i=l}^{m} f_i > 0.$$

with $f_m > 0$ and $f_l > 0$. Then $l > m+1$ (since $f \in \mathbb{R}_a^n$) and $|f_{l+1}| \leq f_l$ since otherwise

$$\left| \sum_{i=l+2}^{m} f_i \right| > \left| \sum_{i=l}^{m} f_i \right| = \|f\|.$$

We also know that $|f_{l+2}| \leq |f_{l+1}|$ because f is not contractable. $|f_{l+2}| \geq |f_{l+3}| \geq \ldots \geq |f_m|$ follows in the same way. We show similarly that $|f_m| \geq |f_l|$. Therefore

$$|f_l| = |f_{l+1}| = \ldots = |f_m| := \beta > 0$$

$m-l$ must be odd and

$$\|f\| = \sum_{i=l}^{m} f_i = \sum_{i=l}^{m} (-1)^{i+1} \cdot \beta = \beta = |f_l|.$$

We have reached the desired contradiction. Lemma 1 follows.

Lemma 2: If $f \in \mathbb{R}_a^n$ and is contractable then $\|f\|=\|\check{f}\|$.

Proof: It is easy to see that $\|f\| \geq \|\check{f}\|$. We assume that $\|f\| < \|\check{f}\|$

and try to find a contradiction. In this case we have

$$\|f\| = \left|\sum_{i=l}^{m} f_i\right|$$

for some l and m where the l and m must satisfy at least one of the following four alternatives: $l=\alpha+1$ (i), $l=\alpha+2$ (ii), $m=\alpha$ (iii), $m=\alpha+1$ (iv), where $\alpha=\alpha(f)$ is defined as above.

Case (i): $l=\alpha+1$. We assume without loss of generality that

$$\sum_{i=l}^{m} f_i > 0.$$

Hence $f_{\alpha+1}\geq 0$. Also $f_{\alpha+2}\leq 0$, $|f_{\alpha+2}|\geq f_{\alpha+1}$, and $m\geq\alpha+3$. Then

$$\|f\| = \sum_{i=\alpha+1}^{m} f_i \leq \sum_{i=\alpha+3}^{m} f_i = \sum_{i=\alpha+3}^{m-2} \check{f}_i \leq \|\check{f}\|.$$

This is the desired contradiction.

The cases (ii) to (iv) lead to similar contradictions. Lemma 2 follows.

We are now in the position to give the steps (and the justification) of an algorithm for computing $\|f\|$:

1. Calculate $h=\hat{f}$.
2. If h is not contractable, then compute $\|f\|=\|h\|$ with the help of lemma 2. Otherwise, replace h by $\check{h}$ and repeat step 2.

The algorithm is also order n^2 and there are certain theoretical worst case examples where it requires more operations than the direct evaluation. However, it functions much faster in practice. In Table 2, a comparison of calculation times of the two algorithms is given. The number of operations needed to calculate $\|\cdot\|$ using directly is only dependent on n. The algorithm is data dependent. Two tests were made, one with vectors of pseudo random numbers uniformly distributed in [-1/2,1/2] and one with laser intensity data with mean value removed. A sample of this data is shown in Figure 2. The algorithm is also practicable with patterns of 4096 points. On an IBM AT, one such calculation on fabric pattern data take less than a minute.

V CONCLUSIONS

We have discussed some measure space distances and how they can be applied to problems in Pattern Recognition. In particular, we have examined the problem of evaluation of irregularities in thickness of a fabric. The discrete irregualrity function based on Discrepancy had suitable properties and can be calculated relatively quickly. The method is currently being used to monitor production quality and to evaluate new production techniques in Industry.

For other applications, particularly those where the number of measured data points is small, the Bounded Lipschitz Distance seems to be very promising.

Almost all of what is discussed here can be extended to patterns of higher dimension. Discrepancy and Bounded Lipschitz Distance can be defined over patterns on squares or in cubes. The only thing which has no obvious extension to higher dimension is the algorithm we described. More sophisticated ideas, maybe using number-theoretic methods, will be necessary.

REFERENCES

[1] Tou, J.T. and Gonzalez, R.C. (1974). Pattern Recognition Principles. Addison-Wesley, Reading, Mass.

[2] Billingsley, P. (1979). Probability and Measure. Wiley, New York.

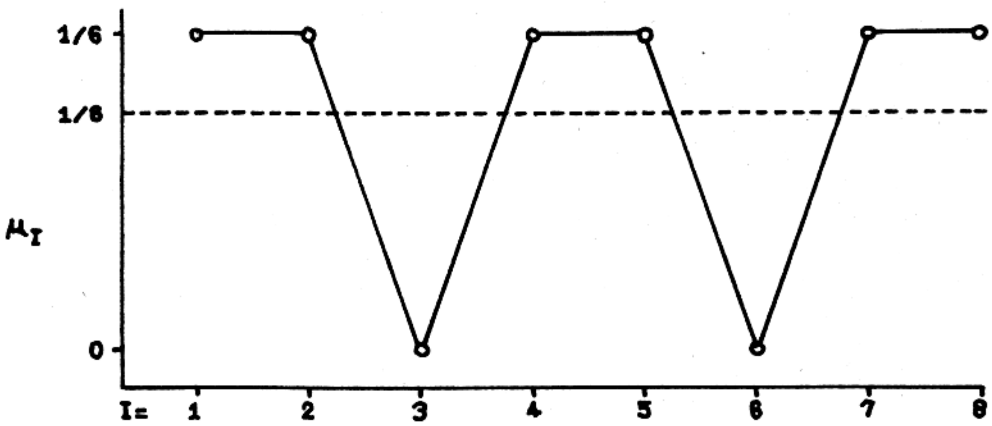
[3] Parthasarathy, K.R. (1967). Probability Measures on Metric Spaces. Academic Press, New York.

[4] Friedrich, V. (1972). Stetige Transportoptimierung. VEB Deutscher Verlag der Wissenschaften, Berlin.

[5] Weyl, H. (1916). Über die Gleichverteilung von Zahlen mod. Eins. Math. Ann. 77, 313-352.

[6] Kuipers, L. and Niederreiter, H. (1974) Uniform Distribution of Sequences. Wiley, New York.

PATTERN A



PATTERN B

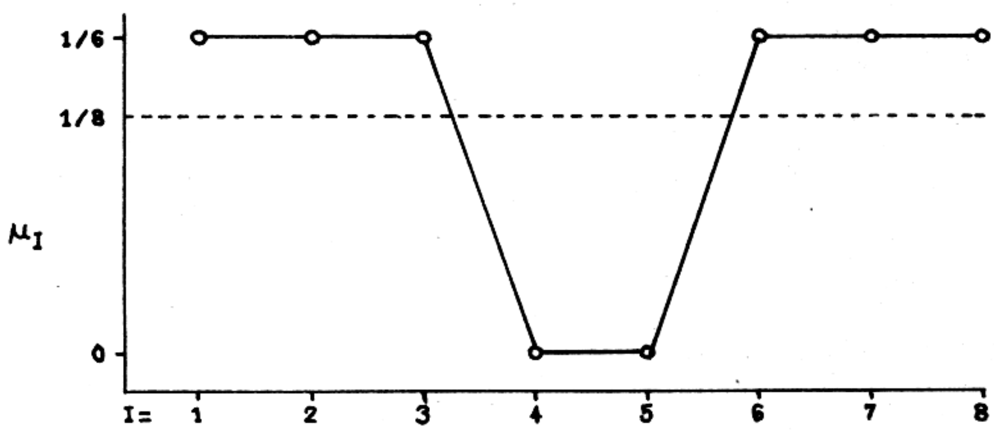


FIGURE 1: TWO TEST PATTERNS (WITH N=8)

	PATTERN A	PATTERN B
I'_V	0.5000	0.5000
I'_B	0.0417	0.0625
I'_*	0.1250	0.2500

TABLE 1: PERFORMANCE OF I'_V, I'_B AND I'_* ON THE TEST PATTERNS IN FIGURE 1

	N=100	N=1000	N=2000
DIRECT CALCULATION	0.120	6.090	21.08
ALGORITHM RANDOM DATA	0.070	0.250	0.450
ALGORITHM LASER INTENSITY DATA	0.004	0.070	0.200

TABLE 2: TIME IN SECONDS FOR 10 NORM CALCULATIONS ON A SIEMENS MAINFRAME COMPUTER

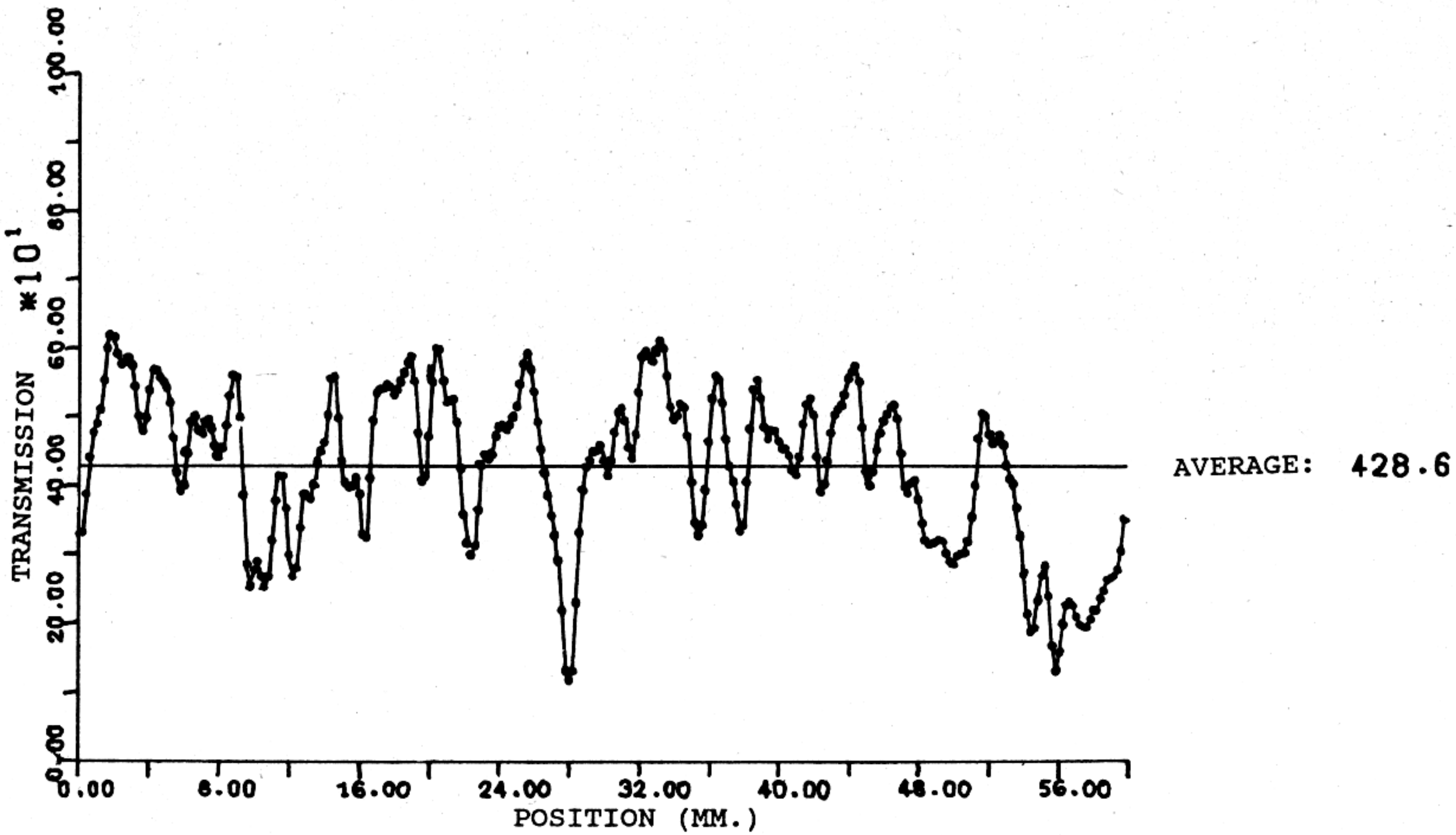


FIGURE 2: LASER TRANSMISSION DATA FOR AN ARTIFICIAL FABRIC

BISHER SIND FOLGENDE BERICHTE ERSCHIENEN:

Nr. 1 FORSCHUNG

W.G. Eschmann und Ralph Götz
Optimierung von Gelenksechsecken

Nr. 2 WEITERBILDUNG

H. Neunzert, M. Schulz-Reese
Mathematische Weiterbildung

Nr. 3 FORSCHUNG

W. Krüger
The Trippstadt Problem

Nr. 4 WEITERBILDUNG

H. Neunzert, M. Schulz-Reese, K.E. Hoffmann
Mathematics in the University and Mathematics in Industry - Complement or Contrast?

Nr. 5 FORSCHUNG

A.K. Louis
The Limited Angle Problem in Computerized Tomography

Nr. 6 FORSCHUNG

W. Krüger
Regression für Ellipsen in achsenparalleler Lage
(Interner Bericht)

Nr. 7 FORSCHUNG

Th. Mietzner
Umströmung von Ecken und Kanten

Nr. 8 FORSCHUNG

W. Krüger, J. Petersen
Simulation und Extrapolation von Rainflow-Matrizen

Nr. 9 FORSCHUNG

W. Krüger, M. Scheutzow u. A. Beste, J. Petersen
Markov- und Rainflow-Rekonstruktionen stochastischer Beanspruchungszeitfunktionen

Nr. 10 FORSCHUNG

Th. Mietzner
Umströmung von Ecken und Kanten, Teil 2

Nr. 11 FORSCHUNG

H. Ploss
Simulationsmethoden zur Lösung der Boltzmanngleichung

Nr. 12 FORSCHUNG

M. Keul
Mathematische Modelle für das Zeitverhalten stochastischer Beanspruchungszeitfunktionen

Nr. 13 AUSBILDUNG

W. Krüger, H. Neunzert, M. Schulz-Reese
Fundamentals of Identification of time series

Nr. 14 FORSCHUNG

H. Moock
Ein Mathematisches Verfahren zur Optimierung von Nocken

Nr. 15 FORSCHUNG

F.-J. Pfreundt
Berechnung und Optimierung des Energiegewinnes bei Anlagen zur Lufterwärmung mittels Erdkanal

Nr. 16 FORSCHUNG

F.-J. Pfreundt
Berechnung einer 2-dimensionalen Kanalströmung mit parallel eingeblasener Luft

Nr. 17 FORSCHUNG

G. Alessandrini
Some remarks on a problem of sound measurements from incomplete data

Nr. 18 AUSBILDUNG

W. Diedrich
Einfluß eines Latentwärmespeichers auf den Wärmefluß durch eine Ziegelwand

Nr. 19 FORSCHUNG

M. Stöhr
Der Kalman-Filter und seine Fehlerprozesse unter besonderer Berücksichtigung der Auswirkung von Modellfehlern

Nr. 20 FORSCHUNG

H. Babovsky
Berechnung des Schalldrucks im Innern eines Quaders

Nr. 21 FORSCHUNG

W. G. Eschmann
Toleranzuntersuchungen für Druckmessgeräte

Nr. 22 FORSCHUNG

G. Schneider
Stratification of solids, a new perspective in three dimensional computer aided design

Nr. 23 FORSCHUNG

H.-G. Stark
Identifikation von Amplituden und Phasensprüngen im Intensitätsverlauf eines Nd-YAG Festkörperlasers

Nr. 24 FORSCHUNG

M. Scheutzow
Einfache Verfahren zur Planung und Auswertung von Navigationsversuchsfahrten

Nr. 25 FORSCHUNG

G.R. Dargahi-Noubary
A Parametric Solution for Simple Stress-Strength Model of Failure with an Application

Nr. 26 FORSCHUNG

U. Helmke, D. Prätzel-Wolters
Stability and Robustness Properties of Universal Adaptive Controllers for First Order Linear Systems

Nr. 27 FORSCHUNG

G. Christmann
Zeitreihen und Modalanalyse

Die Berichte der Arbeitsgruppe Technomathematik können angefordert werden bei:

UNIVERSITÄT KAISERSLAUTERN
Fachbereich Mathematik
Arbeitsgruppe Technomathematik
Postfach 3049

6750 Kaiserslautern