Journal of Global Optimization **11:** 361–376, 1997.

The Chebyshev Hyperplane Optimization Problem

G. STILL and M. STRENG
Department of Applied Mathematics, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

(Received: 19 April 1996; accepted: 20 January 1997)

Abstract. We consider the following problem. Given a finite set of points y^j in $\mathbb{R}^n$ we want to determine a hyperplane H such that the maximum Euclidean distance between H and the points y^j is minimized. This problem (CHOP) is a non-convex optimization problem with a special structure. For example, all local minima can be shown to be strongly unique. We present a genericity analysis of the problem. Two different global optimization approaches are considered for solving (CHOP). The first is a Lipschitz optimization method; the other a cutting plane method for concave optimization. The local structure of the problem is elucidated by analysing the relation between (CHOP) and certain associated linear optimization problems. We report on numerical experiments.

Mathematics Subject Classifications (1991). 90C30, 65D10, 41A50.

Key words: Global optimization, curve fitting, Chebyshev approximation.

1. Introduction

Let $n, m \in \mathbb{N}$ be fixed numbers such that $m \geq n+1$. We use the abbreviation $J = \{1, \cdots, m\}$. In the whole paper $Y = \{y^j \in \mathbb{R}^n \mid j \in J\}$ will be a set of m different points in $\mathbb{R}^n$. As usual, $\| y \|$ will denote the Euclidean norm of $y \in \mathbb{R}^n$. We want to find a hyperplane

$$H = \{y \in \mathbb{R}^n \mid c^T y = \alpha\}, \quad 0 \neq c \in \mathbb{R}^n,\ \alpha \in \mathbb{R},$$

such that the maximum of all Euclidean distances between H and the points in Y is minimized. Since the Euclidean distance between a point y and H is given by $|c^T y - \alpha| / \sqrt{c^T c}$, this problem can be written as

$$\text{(CHOP)} \qquad \min_{\substack{0 \neq c \in \mathbb{R}^n \\ \alpha \in \mathbb{R}}} \max_{j \in J} \frac{|c^T y^j - \alpha|}{\sqrt{c^T c}}$$

and will be called the Chebyshev Hyperplane Optimization Problem (CHOP).

In order to solve (CHOP), we can minimize the numerator subject to a normalization of the denominator. This leads to the following constrained optimization problem:

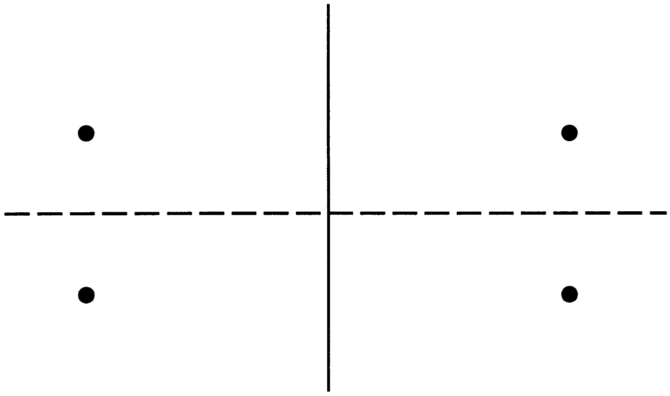

Figure 1. Four points in $I\!R^2$ with a local minimizer of (CHOP) (solid line) and a global minimizer (dashed line).

$$\text{(Q)} \qquad \min f_Q(c,\alpha,r) := r, \quad \text{subject to}$$

$$\begin{cases} q_j^{\pm}(c,\alpha,r) := \pm(c^T y^j - \alpha) - r \le 0, & j \in J \\ c^T c = 1\,. \end{cases}$$

Because of the constraint $c^T c = 1$, (Q) represents a non-convex problem. In particular, (CHOP) can have local solutions which are not global minimizers (see the example for $n = 2, m = 4$, as indicated in Figure 1).

Equivalently, in (CHOP), we can maximize the denominator $\sqrt{c^T c}$ (or minimize $-\sqrt{c^T c}$) subject to a normalization of the numerator. This approach leads to the following concave optimization problem,

$$\text{(P)} \qquad \min f_P(c,\alpha) := -c^T c, \quad \text{subject to}$$
$$p_j^{\pm}(c,\alpha) := \pm(c^T y^j - \alpha) \le 1, \quad j \in J.$$

The feasible set of (Q) resp. (P) will be denoted by Z_Q resp. Z_P.

In the next section it will become clear that (P) and (Q) are equivalent reformulations of (CHOP). We use formulation (P) to demonstrate that any local minimizer of (CHOP) is a strict local minimizer of order 1. In Section 3 we will use (P) for a general position analysis of (CHOP). In Section 4, we propose two methods for the determination of global minimizers of (CHOP), one based on problem (Q), using Lipschitz optimization techniques, and the other based on problem (P), using a cutting plane technique. In the last section we discuss some linear problems related to (CHOP).

Problem (CHOP) is a special case of the problem of approximation of the set Y by a linear (or non-linear) manifold. Another special case, the approximation of Y by a point (also known as the “minimal covering sphere problem” or “Chebyshev center problem”) has a long history (see e.g. [7] [6]). The approximation of a point set by straight lines has been considered in [13, 8, 1]. Investigations on the more general problem can be found in [12] for the Chebyshev norm and in [10] for the least square norm. For complexity aspects of such problems we refer to [2].

2. Optimality Conditions

In the whole paper, we assume that the point set Y satisfies the condition

C1: *Not all points in Y are contained in one hyperplane.*

For later purposes we give some equivalent conditions for $(C1)$ which can be proved by elementary means.

LEMMA 1. *Given $Y = \{y^j | j \in J\}$, the conditions i–v are equivalent.*

i. $(C1)$ is satisfied.

ii. $span \left\{ \binom{y^j}{-1} \;\mid\; j \in J \right\} = \mathbb{R}^{n+1}$.

iii. (Q) has all solutions $(\bar{c}, \bar{\alpha}, \bar{r})$ satisfying $\bar{r} > 0$.

iv. (P) is bounded.

v. The feasible set Z_P of (P) is compact.

For both formulations (P) and (Q) of (CHOP), under condition (C1), the Mangasarian–Fromovitz Constraint Qualification holds, as can easily be seen. So, necessary conditions for local minimizers involve the Kuhn–Tucker condition instead of the more general Fritz John condition. It is not difficult to show that a feasible point (c, α, r) of (Q) satisfying $r > 0$ is a Kuhn–Tucker point of (Q) if and only if the point $(\frac{c}{\|c\|}, \frac{\alpha}{\|c\|})$ is a Kuhn–Tucker point of (P) (with the same active index set $J^* \subset J$ and the same distribution of non-zero multipliers).

It will now be shown that any local minimizer of problem (P) is a strict local minimizer of order 1 (also called a strongly unique minimizer). Due to the correspondence between the Kuhn–Tucker points of (P) and (Q), this is also the case for local minimizers of (Q).

Let (c^*, α^*) be a local minimizer for (P). Then it is called a strict local minimizer of order 1, if there exist a neighbourhood U^* of (c^*, α^*) and a constant $\gamma > 0$ such that for all feasible (c, α) in U^*,

$$- c^T c \geq -(c^*)^T c^* + \gamma \, \|(c, \alpha) - (c^*, \alpha^*)\|. \tag{1}$$

THEOREM 1. *Let (c^*, α^*) be feasible for (P) with active index set J^*. Then, the conditions i–iii are equivalent.*

i. (c^*, α^*) is a local minimum.

ii. (c^*, α^*) is a strict local minimum of order 1.

iii. $\binom{c^*}{0} \in$ int D^*, where $D^* = \left\{ \sum_{j \in J^*} \mu_j \sigma_j \binom{y^j}{-1} \;\mid\; \mu_j \geq 0 \right\}$. Here, $\sigma_j = \pm 1$ if $p_j^{\pm}$ is active at (c^*, α^*).

iv. If $|J^*| = n+1$, then, i–iii are equivalent with: the Kuhn–Tucker condition (2) is valid with all $\mu_j^* > 0$ and $\binom{y^j}{-1}$, $j \in J^*$ linearly independent. (Note that in particular, iii implies that $|J^*| \geq n+1$.)

Proof. i $\Rightarrow$ ii: Suppose (c^*, α^*) is a local minimimum for (P). Since (P) satisfies the Mangasarian–Fromovitz Constraint Qualification, the Kuhn–Tucker condition must be satisfied, i.e. there exist $\sigma_j = \pm 1, \quad \mu_j^* \geq 0,\ j \in J^*$ such that

$$\sum_{j \epsilon J^*} \mu_j^* \sigma_j \begin{pmatrix} y^j \\ -1 \end{pmatrix} = \begin{pmatrix} 2c^* \\ 0 \end{pmatrix}. \tag{2}$$

In order to analyze the second order conditions, we define the cone

$$\begin{aligned} C^* &= \{\xi \in \mathbb{R}^{n+1} \mid Df_P(c^*, \alpha^*)\xi \leq 0,\ Dp_j^{\pm}(c^*, \alpha^*)\xi \leq 0,\ j \in J^*\} \qquad (3) \\ &= \{(\hat{\xi}, \xi_{n+1}) \in \mathbb{R}^{n+1} \mid -2\hat{\xi}^T c^* \leq 0,\ \sigma_j(\hat{\xi}^T y^j - \xi_{n+1}) \leq 0,\ j \in J^*\}. \end{aligned}$$

Defining the Lagrangian

$$\mathcal{L}(c, \alpha, \mu) = -c^T c + \sum_{j \in J^*} \mu_j \sigma_j (c^T y^j - \alpha),$$

then, according to the second order necessary optimality condition for feasible (c^*, α^*) (cf. e.g. [11]), to any $\xi \in C^*$, there exists a multiplier vector $\mu^* \geq 0$ such that (2) holds and

$$\xi^T D^2_{(c,\alpha)} \mathcal{L}(c^*, \alpha^*, \mu^*)\xi = \xi^T \begin{pmatrix} -2I & 0 \\ 0 & 0 \end{pmatrix} \xi = -2\hat{\xi}^T \hat{\xi} \geq 0. \tag{4}$$

This implies $\hat{\xi} = 0$. In view of the last equation in (2), at least one of the σ_j's must be equal to $+1$ and -1 (due to $\mu_j^* \geq 0$). Using (3) gives $-\sigma_j \xi_{n+1} \leq 0,\ j \in J^*$ i.e. $\xi_{n+1} = 0$. Thus, if (c^*, α^*) is optimal then $C^* = \{0\}$. By a well-known theorem (see e.g. [11]) the relation $C^* = \{0\}$ implies that (c^*, α^*) is a strict local minimum of (P) of order 1. For $ii \iff iii$ see e.g. [11] and for iv [3]. □

The problem (CHOP) could be generalized to the approximation of an infinite point set Y. If we suppose, that $Y \subset \mathbb{R}^n$ is compact, the problem (CHOP) represents a semi-infinite problem. Most of the theory remains valid. Note, that Y can be replaced by the set of all extreme points of Y. In the case where Y has infinitely many extreme points, the locally strong uniqueness of a solution need no more remain true. Consider for example the problem in $\mathbb{R}^2$ with Y the unit circle. Then, obviously any line through the origin represents an optimal solution.

3. A General Position Analysis and Stability Aspects

In this section we will investigate, what kind of regularity conditions will be fulfilled for a problem (CHOP) if the problem lies in so-called “general position”. Often, such a study is called a genericity analysis.

For fixed $n, m \in \mathbb{N}, m \geq n+1$, a problem (CHOP) or an equivalent problem (P) can be seen as an element from $P_{n,m}$,

$$P_{n,m} = \{y^j \in \mathbb{R}^n \ \mid \ j = 1, \cdots, m\}. \tag{5}$$

This set $P_{n,m}$ can be identified with the (Euclidean) $\mathbb{R}^{nm}$. The following analysis will be done for the problem (P). Putting

$$A = \begin{pmatrix} a^1 \\ a^2 \\ \vdots \\ a^m \end{pmatrix} \text{ where } a^j = \begin{pmatrix} y^j \\ -1 \end{pmatrix}^T \quad \text{and} \quad 1_k = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^k,$$

the feasible set Z_P of (P) becomes

$$Z_P = \left\{ z = \begin{pmatrix} c \\ \alpha \end{pmatrix} \in \mathbb{R}^{n+1} | \begin{pmatrix} A \\ -A \end{pmatrix} z \leq 1_{2m} \right\}. \tag{6}$$

Obviously, Z_P is a polyhedron having inner points (i.e. a polyhedron of full dimension). A feasible point $\bar{z}$ is called a vertex of Z_P if $\bar{z}$ is given as the solution of

$$A_{J_0} z = 1_{n+1} \quad , \tag{7}$$

where J_0 denotes an index set

$$J_0 = \{j_k, \ k = 1, \cdots, n+1 \mid 1 \leq j_1 < \cdots < j_{n+1} \leq 2m\}, \tag{8}$$

satisfying $j \in J_0, \ 1 \leq j \leq m \Longleftrightarrow j + m \notin J_0$, and where A_{J_0} is the matrix

$$A_{J_0} = \begin{pmatrix} \sigma_{j_1} a^{j_1} \\ \vdots \\ \sigma_{j_{n+1}} a^{j_{n+1}} \end{pmatrix}, \quad a^j = a^{|j-m|} \text{ for } j > m, \tag{9}$$

$$\sigma_j = \begin{cases} 1 & \text{if } \ 1 \leq j \leq m \\ -1 & \text{if } \ m+1 \leq j \leq 2m \end{cases}$$

such that A_{J_0} is regular. The vertex $\bar{z} \in Z_P$ is called non-degenerate, if

$$\sigma_j a^j \ \bar{z} < 1 \ \text{ for all } \ j \in \{1, \ldots, 2m\} \backslash J_0. \tag{10}$$

The choice $J_0 \subset \{1, \ldots, m\}$ or $J_0 \subset \{m+1, \ldots, 2m\}$, i.e. all σ_j have the same sign, leads to the vertex $v = \pm e_{n+1}$, which is always degenerate (if $m > n+1$). Note, that the vertices $\pm e_{n+1}$ cannot be optimal. So, we can assume that in (9) at least one of the σ_k's is positive and at least one negative. In the sequel we will use the following result from Stratification Theory (for a proof we refer to [5]).

THEOREM 2. *Let be given a polynomial function* $p: \mathbb{R}^N \to \mathbb{R},\ p \neq 0$. *Then, the set* $p^{-1}(0) = \{x \in \mathbb{R}^N \mid p(x) = 0\}$ *is a closed set of Lebesgue measure zero. (Notation:* $\mu(p^{-1}(0)) = 0$.*)*

We emphasize, that with $z = (c, \alpha)$, also $-z$ is a vertex of (P). Hence, in the following, we tacitly identify z and $-z$ (i.e. two vertices z and $\hat{z}$ are called different iff $z \neq \hat{z}$ and $z \neq -\hat{z}$). We now state the genericity result.

THEOREM 3. *Let* $n, m \in \mathbb{N}$ *be fixed,* $m \geq n + 1$. *Then, the problem set* $P_{n,m}$ *contains an open, dense subset* M_0 *with* $\mu(P_{n,m} \backslash M_0) = 0$ *such that for all problems (P) in* M_0 *the following hold:*
The condition (C1) is satisfied (and by Lemma 1 Z_P *is a compact polyhedron given by the convex hull of its vertices). All vertices* $\bar{z} = (\bar{c}, \bar{\alpha})$ *of* Z_P *(except the vertices* $\pm e_{n+1}$*) are non-degenerate and have different values* $-\bar{c}^T\bar{c}$. *In particular, (P) has a unique global minimum* $z^* = (c^*, \alpha^*)$.

(Note, that by Theorem 1 this solution and all other possible local minimizers are strict minima of order1 and characterized by the conditions in Theorem 1 iv.)

Proof. We firstly show that generically all vertices of Z_P are non-degenerate. Let J_0 be a fixed index set (8). Since $m \geq n+1$, such an index set exists. Consider the condition rank $A_{J_0} = n + 1$ or equivalently $p(y^{j_1}, \ldots, y^{j_{n+1}}) := \det A_{J_0} \neq 0$. Obviously p is a polynomial function, $p \neq 0$. Thus, by Theorem 2 the set M_{J_0} defined by

$$M_{J_0} = \{(P) \in P_{n,m} \mid \operatorname{rank} A_{J_0} = n + 1\},$$

is open, dense in $P_{n,m} (\equiv \mathbb{R}^{nm})$ with $\mu(P_{n,m} \backslash M_{J_0}) = 0$. Now, let J_0 be given as above such that A_{J_0} is regular, and choose $j_0 \notin J_0$. Then for the solution $\bar{z}$ of $A_{J_0}\bar{z} = 1_{n+1}$ we have:

$$a^{j_0}\bar{z} \neq 1 \iff A_{J_0,j_0} := \begin{pmatrix} A_{J_0} & 1_{n+1} \\ a^{j_0} & 1 \end{pmatrix} \text{ has full rank } n+2. \tag{11}$$

As above, by using Theorem 2 it follows, that the set (with both, J_0 and $j_0 \notin J_0$ fixed)

$$M_{J_0,j_0} = \{(P) \in P_{n,m} \mid \operatorname{rank} A_{J_0,j_0} = n + 2\}$$

is open, dense and $\mu(P_{n,m} \backslash M_{J_0,j_0}) = 0$. Now we use the fact that the union of finitely many sets of measure zero has measure zero and that there are only finitely many combinations of basic index sets J_0 as in (8) and $j_0 \notin J_0$. By construction, the intersection of all these sets M_{J_0}, M_{J_0,j_0} only contain problems (P) with nondegenerate vertices. Hence there exists an open and dense subset of $P_{n,m}$ such that for all (P) from this subset all vertices ($\neq \pm e_{n+1}$) are nondegenerate. Note, that by Lemma 1-ii, for any (P)$\in M_{J_0}$ the condition (C1) is valid. Thus, by Lemma 1, the feasible set Z_P is a compact polyhedron. It is well-known that any compact polyhedron can be described as the convex hull of its vertices (in particular Z_P actually has vertices).

We now prove the conclusion that there is an open and dense subset $M_0 \in P_{n,m}$ such that moreover, all vertices $\bar{z} = (\bar{c}, \bar{\alpha})$ of (P)$\in M_0$ have different function-values. Let us assume, that $\bar{z} = (\bar{c}, \bar{\alpha})$ and $\hat{z} = (\hat{c}, \hat{\alpha})$ are two different (nondegenerate) vertices of Z_P, i.e. with sets $\bar{J}, \hat{J}$ as in (8), $\bar{J} \neq \hat{J}$, we have

$$\bar{z} = A_{\bar{J}}^{-1} 1_{n+1} \ , \ \hat{z} = A_{\hat{J}}^{-1} 1_{n+1} \quad (\text{ and } \bar{z} \neq \pm \hat{z}). \tag{12}$$

With the adjoint $A_{\bar{J}}^{ad}$ of $A_{\bar{J}}$ we can write $A_{\bar{J}}^{-1} = \frac{1}{detA_{\bar{J}}} A_{\bar{J}}^{ad}$ and accordingly $A_{\hat{J}}^{-1} = \frac{1}{detA_{\hat{J}}} A_{\hat{J}}^{ad}$. Let $\bar{B} \ \ (\hat{B})$ denote the $n \times (n+1)$-matrix obtained by deleting the $(n+1)^{th}$ row from $A_{\bar{J}}^{ad} \ \ (A_{\hat{J}}^{ad})$. Then $\bar{c} = \frac{1}{detA_{\bar{J}}} \bar{B} 1_{n+1}$ and $\hat{c} = \frac{1}{detA_{\hat{J}}} \hat{B} 1_{n+1}$. Hence, $\bar{z}, \hat{z}$ have the same value $-\bar{c}^T \bar{c} = -\hat{c}^T \hat{c}$ iff

$$p(A_{\hat{J}}, A_{\bar{J}}) := (detA_{\hat{J}})^2 1_{n+1}^T \bar{B}^T \bar{B} 1_{n+1} - (detA_{\bar{J}})^2 1_{n+1}^T \hat{B}^T \hat{B} 1_{n+1} = 0.$$

This relation represents a polynomial equation with a non-vanishing polynomial p depending on the variables $y^j, j \in \bar{J} \cup \hat{J}$ and from Theorem 2 it follows, that the set $p^{-1}(0)$ is a closed set $S_0 \in P_{n,m}$ of measure zero. Thus $M_3 = P_{n,m} \backslash S_0$ is open, dense and $\mu(P_{n,m} \backslash M_3) = \mu(S_0) = 0$. By construction, for $(P) \in M_3$, the vertices $\bar{z}, \hat{z}$ given by $\bar{J}, \hat{J}$ have different values. Since there are only finitely many such index sets $\bar{J}, \hat{J}; \bar{J} \neq \hat{J}$ the intersection M_0 of all relevant sets satisfies the conditions of Theorem 3. $\square$

It is obvious that by using the results on the relations between the problems (Q) and (P) in Section 2, corresponding genericity statements can be formulated for problem (Q).

We finish this section with some remarks on stability. Let be given a problem $\overline{Y} = \{\overline{y}^1, \ldots, \overline{y}^m\} \in P_{n,m}$ and a local minimizer $\bar{z} = (\bar{c}, \bar{\alpha})$ of $(\overline{P}) = P(\overline{Y})$. ($P(Y)$ will denote the problem (P) in dependence on the point set Y.) We will discuss the question whether this solution $\bar{z}$ persists under small perturbations of $\overline{Y}$. For $n = 2$, this stability question has been considered in [12]. For the case that $\bar{z}$ is a non-degenerate vertex of $Z_{P(\overline{Y})}$ (cf. 10) the following result holds:

Strong Stability Result: Suppose that $\bar{z}$ is a local solution of $P(\overline{Y})$ such that $\bar{z}$ is a non-degenerate vertex of $Z_{P(\overline{Y})}$. Then, there exist neighbourhoods U of $\bar{z}$ and V of $\overline{Y}$ and a function $z: \ V \to \ U, \ z(\overline{Y}) = \bar{z}$ such that for all $Y \in V$ the vector $z(Y) \in U$ is the unique local minimizer (strict of order 1) of $P(Y)$ in U. Moreover, the solution function z is (infinitely many times) continuously differentiable.

This result follows directly by considering locally around $Y = \overline{Y}$ the equation

E(Y): $$\bar{\sigma}_j (c^T y^j - \alpha) = 1, \qquad j \in J^*$$

and (2) for a solution $z = (c, \alpha)$ of $P(Y)$, with corresponding multipliers μ_j. When $\bar{z}$ is a non-degenerate vertex, we have $|J^*| = n + 1$ and $\bar{z} = (\bar{c}, \bar{\alpha})$ is the unique solution of the equation $E(\overline{Y})$. This follows by Theorem 1 iv. By continuity there exists a neighbourhood V of $\overline{Y}$ such that for all $Y \in V$ there is a unique solution $z(Y)$ of $E(Y)$.

When at a local minimum $\bar{z}$ of $P(\overline{Y})$ we have $|J^*| > n+1$, then, a small perturbation of $\overline{Y}$ might result into a 'bifurcation' into several local minima near $\bar{z}$. Take as a simple example the problem set $\overline{Y} = \{\binom{-2}{1}, \binom{-2}{-1}, \binom{2}{1}, \binom{2}{-1}\}$ with (global) solution $\bar{z} = (0, 1, 0)$. Consider a small perturbation $Y_\epsilon = \{\binom{-2}{1}, \binom{-2}{-1}, \binom{2}{1-\epsilon}, \binom{2}{-1+\epsilon}\}$. Then for any $\epsilon > 0$ the problem $P(Y_\epsilon)$ has two (global) minima $z^1 = (\epsilon/4, 1, -\epsilon/2)$, $z^2 = (-\epsilon/4, 1, \epsilon/2)$.

However he local minima cannot disappear completely by a small perturbation. This is stated in the following

Weak Stability Result: Suppose, $\bar{z}$ is a local solution of $P(\overline{Y})$. Then, there exists a neighbourhood V of $\overline{Y}$ such that for any $Y \in V$ there is at least one local solution $z(Y)$ of $P(Y)$. Moreover with a constant $\kappa > 0$ we have $||z(Y) - \bar{z}|| \le \kappa ||Y - \overline{Y}||$.

This result follows by a well-known (weak) stability result valid for non-linear optimization problems at a strict local solution $\bar{z}$ of order 2 under Mangasarian–Fromovitz Constraint Qualification (cf. e.g.[9]).

Since for the generic set $M_0 \subset P_{n,m}$ of Theorem 3 in particular all local minima of problems $(P) \in M_0$ are non-degenerate vertices of Z_P, the genericity result in Theorem 3 together with the Strong Stability Result leads to

THEOREM 4. *For all $(P) \in M_0$ all local minima $\bar{z}$ of (P) are strongly stable (in the sense ot the Strong Stability Result).*

4. Two Methods for Solving CHOP

In this section we will briefly discuss two different methods of global optimization to solve problem (CHOP).

The first method is a so-called Lipschitzian optimization approach. By defining $S^{n-1} = \{c \in \mathbb{R}^n | c^T c = 1\}$, (CHOP) can be written as $\min_{c \in S^{n-1}} \min_{\alpha \in \mathbb{R}} \max_{j \in J} |c^T y^j - \alpha|$. For fixed $c \in S^{n-1}$ we consider the function

$$F(c) := \min_{\alpha \in \mathbb{R}} \max_{j \in J} |c^T y^j - \alpha|.$$

Then, obviously (CHOP) is equivalent to the problem

$$(\tilde{Q}) \qquad \min_{c \in S^{n-1}} F(c). \tag{13}$$

Putting $M(c) = \max_{j \in J} c^T y^j, \;\; m(c) = \min_{j \epsilon J} c^T y^j$ we can write

$$F(c) = \frac{M(c) - m(c)}{2}. \tag{14}$$

This function F is Lipschitz continuous,

$$|F(c_1) - F(c_2)| \leq \lambda \parallel c_1 - c_2 \parallel \quad \text{for all } c_1, c_2 \in \mathbb{R}^n,$$

with Lipschitz constant $\lambda = \max_{j \epsilon J} \parallel y^j \parallel$. The easy proof is omitted.

The Lipschitzian method to solve $(\tilde{Q})$ is based on the following simple idea: Let $K \subset \mathbb{R}^n$ be compact. We denote the diameter of K by $d(K)$, $d(K) = \max_{c_1, c_2 \epsilon K} \parallel c_1 - c_2 \parallel$. Then, for given $\bar{c} \in K$, by using the Lipschitz condition for F, we find for any $c \in K$ the inequality $F(\bar{c}) \leq F(c) + |F(\bar{c}) - F(c)| \leq F(c) + \lambda d(K)$. Consequently

$$\min_{c \epsilon K} F(c) \geq F(\bar{c}) - \lambda d(K). \tag{15}$$

In every step of the following branch-and-bound algorithm A1 this inequality (15) enables us to discard certain parts of the feasible set S^{n-1} of $(\tilde{Q})$ from a further minimum search. (See [4, pp. 111–140], and [12] for more details.)

A1: Algorithm for solving $(\tilde{Q})$ (see (13))
Start: Choose $\epsilon > 0$ and an initial partition S_0 of S^{n-1}, $S_0 = \cup_{\mu=1}^{m_0} S_{0,\mu}$. Compute $\lambda = \max_{j \epsilon J} \parallel y^j \parallel$.

Step $k \to k+1$: Given a partition of a set $S_k \subset S^{n-1}$, $S_k = \cup_{\mu=1}^{m_k} S_{k,\mu}$ such that $\min_{c \epsilon S^{n-1}} F(c) = \min_{c \epsilon S_k} F(c)$, we proceed as follows.

1. For $\mu = 1, \cdots, m_k$; Choose a point $c_{k\mu} \epsilon S_{k,\mu}$ and compute $\alpha_{k\mu} = F(c_{k\mu})$, $\beta_{k\mu} = \alpha_{k\mu} - \lambda d(S_{k,\mu})$.
2. Compute $\alpha_k = \min_{\mu=1,\cdots,m_k} \alpha_{k\mu}$, $\beta_k = \min_{\mu=1,\cdots,m_k} \beta_{k\mu}$. If $\alpha_k - \beta_k < \varepsilon$ stop with an approximate minimum $c_{k\bar{\mu}}$ such that $F(c_{k\bar{\mu}}) = \alpha_k$.
3. Delete from S_k all sets $S_{k,\mu}$ such that $\alpha_k < \beta_{k\mu}$. Choose a finer partition S_{k+1} of the remaining sets, $S_{k+1} = \cup_{\mu=1}^{m_{k+1}} S_{k+1,\mu}$.

Another approach to solve (CHOP) is to apply methods from concave optimization to the equivalent problem (P). With the notations of Section 3, the problem (P) consists of minimizing the concave function $f(c, \alpha) = -c^T c$ on the polyhedron Z_P (cf. (6)). Due to condition $(C1)$ the feasible set $Z = Z_P$ is compact (cf. Lemma 1). It is well-known that the global minimum of a concave function f over a compact polyhedron Z is always attained at some vertex of Z. (See [4, p. 10] for a proof.) This fact allows to apply the Simplex method. We briefly outline a Simplex method combined with a cutting plane method for solving problem (P). For more details the reader is referred to [4].

The method is based on the following construction, which is applicable to any problem of minimizing a concave function on a polyhedral set. Suppose z^0 is a nondegenerate vertex of Z_P defined by $A_{J_0} z^0 = 1_{n+1}$ (cf. Section 3). Let γ be given, such that $f(z^0) \geq \gamma$. Define for $j = 1, \cdots, n+1$ (with e_j the unit vectors in $\mathbb{R}^{n+1}$)

$$\tau_j = sup\{t \geq 0 \mid f(z^0 - tA_{J_0}^{-1} e_j) \geq \gamma\} \text{ and } v^j = z^0 - \tau_j A_{J_0}^{-1} e_j. \tag{16}$$

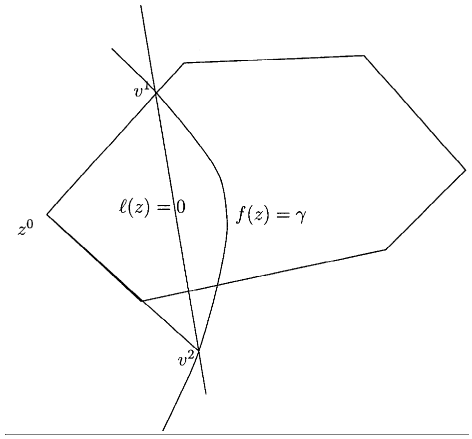

Figure 2. Illustration of the cutting plane method.

Note, that the vectors $-A_{J_0}^{-1}e_j$ give the directions of the $n+1$ edges emanating from z^0. We suppose that $0 < \tau_j < \infty,\ j = 1, \cdots, n+1$. Then, we can define $q = (\frac{1}{\tau_1}, \cdots, \frac{1}{\tau_{n+1}})^T$ and the linear function

$$\ell(z) = -q^T A_{J_0} z + q^T 1_{n+1} - 1. \tag{17}$$

This function satisfies $\ell(z^0) = -1$ and (cf. (16)) $\ell(v^j) = 0,\ j = 1, \cdots, n+1$. Consequently, the equation $\ell(z) = 0$ defines the hyperplane going through the points $v^j,\ j = 1, \cdots, n+1$ (cf. Figure 2).

Let S_0 denote the $(n+1)$-simplex with vertices $z^0, v^1, \cdots, v^{n+1}$. By construction, $f(z^0) \geq \gamma$ and $f(v^j) \geq \gamma,\ j = 1, \cdots, n+1$. The above mentioned fact that the convave function f attains its minimum at a vertex of S_0 implies that $\min_{z \epsilon S_0} f(z) \geq \gamma$. Since $\{z \in Z_P \mid \ell(z) \leq 0\} \subset S_0$ the linear inequality $\ell(z) \geq 0$ defines a so-called 'γ-valid cut', i.e.

$$Z_\gamma = \{z \in Z_P \mid f(z) < \gamma\} \subset \{z \in Z_P \mid \ell(z) > 0\}. \tag{18}$$

In particular, if the set $\{z \in Z_P \mid \ell(z) > 0\}$ is empty, then $\gamma \leq \min_{z \in Z_P} f(z)$.

A2: Simplex-cutting-plane algorithm to solve (P).

Start: Find a vertex $z^0 \in Z_P$. Put $k = 0, \gamma = f(z^0)$, $z_{opt} = z^0$ and $Z_0 = Z_P$.

step $k \to k+1$: Given γ and a vertex z^k of a polyhedron $Z_k \subset Z_P$ such that $\min_{z \in Z_k} f(z) = \min_{z \in Z_P} f(z)$ we proceed with:

1. If z^k is a local minimizer of $\min_{z \epsilon Z_k} f(z)$, then goto 3.
2. Find a neighbouring vertex $\bar{v}$ of z^k in Z_k such that $f(\bar{v}) < f(z^k)$. Put $z^k = \bar{v}$ and goto 1.

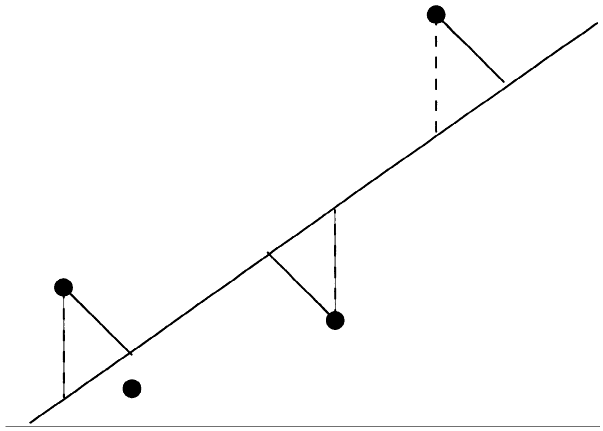

Figure 3. Illustration of a linear problem related to (CHOP).

3. If $\gamma > f(z^k)$, put $\gamma = f(z^k)$ and $z_{opt} = z^k$. Compute from z^k a 'γ-valid cut' given by $\ell_k(z) \geq 0$ (see 18 and the construction above). Compute a vertex solution $\hat{v}$ of the linear program
   max $\ell_k(z)$ s.t. $z \in Z_k$.
   If $\ell_k(\hat{v}) \leq 0$, then stop with a global solution z_{opt}. Otherwise, put $z^{k+1} = \hat{v}$ and $Z_{k+1} = \{z \epsilon Z_k | \ell_k(z) \geq 0\}$.

For a discussion of the convergence of algorithms of the type A2 we refer to [4, pp. 99, 183].

REMARK 1. For problem (P) it is evident that if (c, α) is feasible, also $-(c, \alpha)$ is feasible. Therefore, we can assume a further restriction for $z = (c, \alpha)$ such as $c_n \geq 0$. In view of the genericity results of Section 3, we can expect that, in general, all vertices of Z_P (apart from the vertices $\bar{z} = \pm e_{n+1}$) will be non-degenerate.

5. Linear Problems Related to CHOP

We consider again the problem (CHOP) in the form (Q). Take for example the problem (Q) in $\mathbb{R}^2$. A hyperplane H is given by the equation $c^T y - \alpha = c_1 y_1 + c_2 y_2 - \alpha = 0$ with $c_1^2 + c_2^2 = 1$. By assuming $c_2 \neq 0$, this relation can be written as

$$y_2 = h(y_1) = -\frac{c_1}{c_2} y_1 + \frac{\alpha}{c_2}.$$

Geometrically, (Q) is the problem of finding a linear function $h(y_1) = \gamma y_1 - \beta$ such that the maximal Euclidean distance in $\mathbb{R}^2$ between the points $(y_1^j, y_2^j) \in Y$ and the graph of h is minimized. By minimizing simply the maximal distance $|y_2^j - h(y_1^j)|$ we arrive at a corresponding linear approximation problem (see Figure 3).

We now will generalize this idea to problems (Q) in $\mathbb{R}^n$ and discuss the relation between the non-convex problem (Q) and the associated linear problems.

Let $k,\ 1 \le k \le n$, be fixed. Suppose, (c, α, r) is feasible for (Q) such that $c_k \neq 0$. Then, we define

$$\gamma_k = -\frac{1}{c_k}(c_1, \dots, c_{k-1}, c_{k+1}, \dots, c_n), \quad \alpha_k = -\frac{\alpha}{c_k}, \quad r_k = \frac{r}{|c_k|} \tag{19}$$

and $t_k^j = (y_1^j, \dots, y_{k-1}^j, y_{k+1}^j, \dots, y_n^j),\ j \in J$. We emphasize, that here, γ_k and t_k^j denote vectors in $\mathbb{R}^{n-1}$. Now, the vector $(\gamma_k, \alpha_k, r_k)$ is a feasible solution of the following linear optimization problem:

$$(\mathrm{Q}_k) \qquad \min\ r_k \ \text{ s.t. } \pm(\gamma_k^T t_k^j - \alpha_k - y_k^j) - r_k \le 0,\ \ j \in J.$$

In comparison with (Q), the problem (Q_k) represents the approximation of the data $\{t_k^j, y_k^j\}_{j\in J}$ by the linear function $h(t) = \gamma^T t - \alpha\ \ (\gamma, t \in \mathbb{R}^{n-1},\ \alpha \in \mathbb{R})$ in the Chebyshev norm (cf. Figure 3 for a geometrical illustration).

Similarly, for a feasible point $(\gamma_k, \alpha_k, r_k)$ of (Q_k) by defining $\hat{c} = ((\gamma_k)_1, \dots, (\gamma_k)_{k-1}, -1, (\gamma_k)_k, \dots, (\gamma_k)_{n-1})$ ($(\gamma_k)_j$ denoting the components of γ_k) and

$$c = \frac{\hat{c}}{\sqrt{\hat{c}^T\hat{c}}}, \quad \alpha = \frac{\alpha_k}{\sqrt{\hat{c}^T\hat{c}}}, \quad r = \frac{r_k}{\sqrt{\hat{c}^T\hat{c}}}, \tag{20}$$

the vector (c, α, r) is feasible for (Q). We summarize these facts in the following lemma.

LEMMA 2. *Let (c, α, r) and $(\gamma_k, \alpha_k, r_k)$ correspond according to (19) and (20). Then, (c, α, r) satisfying $c_k \neq 0$ is feasible for (Q) with an active index set $\bar{J}$ if and only if $(\gamma_k, \alpha_k, r_k)$ is feasible for (Q_k) with the same active set $\bar{J}$.*

Although the original problem (Q) is non-convex and the associated programs $Q_k,\ k = 1, \cdots, n$ are linear, Figure 3 indicates that these problems are closely connected. We give two examples to clarify, what kind of situations are possible. Firstly, an example of solutions of (Q_2) (or (Q_1)) which are local but not global minimizers of (Q).

EXAMPLE 1. Let us consider the problem in $\mathbb{R}^2$ given by the points $y^j = (j, j + (-1)^j\mu),\ \ j = 1, 2, 3$, where μ is a parameter, $\mu \in (0, \infty)$. Then, the following holds with the lines $\bar{h}$ and h defined by $\bar{h}(t) = t$, $h(t) = (1 + 2\mu)t - 5\mu$ and with the values $\mu_0 = \sqrt{\frac{1}{2}}$, $\mu_1 = (\sqrt{7} - 1)/2$ (cf. Figure 4): $\bar{h}$ is always the (unique) solution of (Q_2) for all $\mu \ge 0$, and

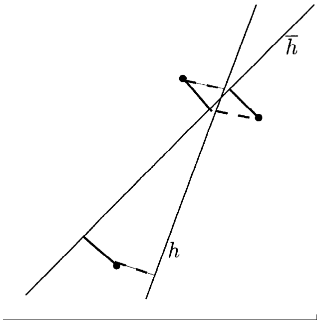

Figure 4. Illustration corresponding to Example 1 for the value $\mu = \mu_1$.

For	$0 < \mu < \frac{1}{2}$:	$\bar{h}$ is the solution of (Q_1), (Q).
For	$\frac{1}{2} < \mu < \mu_0$:	$\bar{h}$ is the solution of (Q) and h is the solution of (Q_1).
For	$\mu_0 < \mu < \mu_1$:	$\bar{h}$ is the solution of (Q); h is a local solution of (Q) and the solution of (Q_1).
For	$\mu_1 < \mu < 1$:	h is the solution of (Q) and (Q_1); $\bar{h}$ is a local solution of (Q).
For	$\mu > 1$:	h is the solution of (Q) and (Q_1).

Next we give a problem, where the solutions of $(Q), (Q_1)$ and (Q_2) are all different.

EXAMPLE 2. We choose the points

$$y^j = \left(cos \left(\frac{\pi}{12} + (j-1)\frac{2}{3}\pi \right), sin \left(\frac{\pi}{12} + (j-1)\frac{2}{3}\pi \right) \right), \quad j = 1, 2, 3,$$

on the unit circle (cf. Figure 5). Then, h_1 is the unique solution of (Q_1), h_2 the unique solution of (Q_2) and (Q) has the three global solutions h_1, h_2, h. The solutions of $(Q_1), (Q_2), (Q)$ depend continuously on small perturbations of the points y^j, $j = 1, 2, 3$. Therefore, it is clear from Figure 5 that by only perturbing y^2 by $\hat{y}^2 = y^2(1 - \epsilon)$ ($\epsilon > 0$ small) we will obtain a unique solution $\hat{h}$ of (Q) parallel to h at a distance $\frac{\epsilon}{2}$. Whereas the solutions of (Q_1) (resp. (Q_2)) will remain near h_1 (resp. (h_2)).

In the situation however, where all problems (Q_k), $k = 1, \ldots, n$, have the same optimal hyperplane, the corresponding feasible point (c, α, r) is the global solution of (Q).

THEOREM 5. *Suppose, $(\bar{c}, \bar{\alpha}, \bar{r})$ is such that the corresponding vectors $(\bar{\gamma}_k, \bar{\alpha}_k, \bar{r}_k)$ (cf. (19)) are solutions of (Q_k) for all $k = 1, \ldots, n$. Then, $(\bar{c}, \bar{\alpha}, \bar{r})$ is the global solution of (Q).*

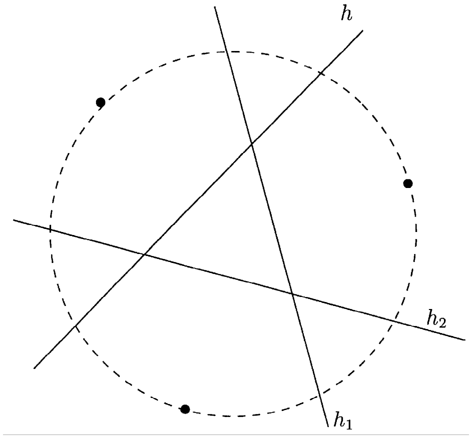

Figure 5. Illustration corresponding to Example 2.

Proof. Suppose, (c, α, r) is feasible for (Q) with $r < \bar{r}$. This leads to a contradiction as follows. It is geometrically clear that the relation $r < \bar{r}$ implies $(c, \alpha) \neq \pm(\bar{c}, \bar{\alpha})$ and also $c \neq \pm\bar{c}$. Since $1 = \sum_{j=1}^{n} \bar{c}_j^2 = \sum_{j=1}^{n} c_j^2$, there must be some index $\tau,\ 1 \le \tau \le n$ such that

$$c_\tau \neq \bar{c}_\tau \qquad \text{and} \qquad |c_\tau| \ge |\bar{c}_\tau|. \tag{21}$$

Defining the feasible solution $(\gamma_\tau, \alpha_\tau, r_\tau)$ of (Q_τ) corresponding to (c, α, r) (cf. (19)) we find

$$r_\tau = \frac{r}{|c_\tau|} < \frac{\bar{r}}{|\bar{c}_\tau|} = \bar{r}_\tau$$

contradicting the optimality of $(\bar{\gamma}_\tau, \bar{\alpha}_\tau, \bar{r}_\tau)$ for (Q_τ). □

6. Numerical Experiments

We finally report on some numerical experiments. The point sets $Y \subset \mathbb{R}^n$ have been generated randomly as follows. Firstly, m points $\hat{y}^j$ are generated randomly in the hyperrectangle $[-100, 100]^{n-1} \times [-d, d]$ (with fixed $d > 0$). Then, we generate randomly c ($c \in \mathbb{R}^n$, $\|c\| = 1$) and α ($\alpha \in \mathbb{R}$). Finally we choose an orthogonal matrix Q which transforms the unit vector e_n into c (i.e. $c = Qe_n$) and apply the affine transformation

$$y^j = Q\hat{y}^j + \alpha c.$$

The problem points are then lying in a corresponding hyperrectangle around the hyperplane $H = \{y \in \mathbb{R}^n \mid y^T c - \alpha = 0\}$. The smaller d (compared with 100) the better the data Y fit the hyperplane H. Firstly we compare the Lipschitz method in

Table I. Comparison between the algorithms A1 and A2 for $n = 3$, $d = 1$ and different m.

m	A2 number of γ-valid cuts	A2 evaluation time in seconds	A1 evaluation time in seconds
10	1	0.05	10.16
30	1	0.10	14.66
50	2	0.27	15.04
100	2	0.49	92.49
200	2	0.87	108.58

algorithm A1 with the cutting plane algorithm A2. Table I contains some typical examples for $n = 3$. The efficiency of the methods is roughly indicated by giving an evaluation time (on the PC used for these examples). For the cutting plane algorithm we also give the number of γ-valid cuts performed in part 3 of A2. In A1 we have chosen a representation of the unit sphere S^2 in spherical coordinates $(r, \varphi, \psi) \in \{1\} \times [-\pi, \pi] \times [0, \frac{\pi}{2}]$. The partitions of S^2 are given by a uniform rectangular grid on $[-\pi, \pi] \times [0, \frac{\pi}{2}]$ by halving the gridlength in every refinement step 3 of A1. We have always taken $\varepsilon = 10^{-10}$ in A1.

From Table I we might conclude that the cutting plane algorithm is much more efficient than the Lipschitz method. The number of γ-valid cuts and the computer time needed in algorithm A2 increases with the number d. This is shown by the results in Table II containing the computer time to run A2 for $m = 100$ and different n and d.

Table II. Computer time for A2 for $m = 100$ and different n and d.

d	n	number of γ-valid cuts	evaluation time in seconds
1	4	2	0.60
1	5	2	1.26
1	10	2	4.83
1	20	3	32.13
15	4	2	0.82
15	5	6	2.80
15	10	53	94.14
50	3	12	1.97
50	4	39	11.25
50	5	127	63.93

We finally investigate numerically the relation between the problem (Q) and the associated linear problems (Q_k), $k = 1, \ldots, n$. Table III contains some results for $m = 20$. For different n and d we have randomly generated 10 problems (Q) and indicate with the vector $(a_0, \ldots, a_n)$ how many of the solutions of the corresponding n problems (Q$_k$) coincided with the solution of (Q). For $n = 3$ for example, $(1, 4, 3, 2)$ means that out of the 10 problems, 1 had none of the solutions of the (Q$_k$)'s common with the solution of (Q) , 4 problems had one of the solutions of the (Q$_k$)'s common with the solution of (Q) etc. and for 2 problems all three solutions of the (Q$_k$)'s coincided with the solution of (Q).

Table III. Relation between (Q) and problems (Q$_k$) for $m = 20$ and different n and d.

n	d	(a_0, a_1, a_2, a_3)	n	d	$(a_0, a_1, a_2, a_3, a_4)$
3	1	(0,0,1,9)	4	1	(0,0,0,2,8)
3	10	(0,3,4,3)	4	10	(2,3,4,1,0)
3	40	(7,3,0,0)	4	40	(5,5,0,0,0)

Table III indicates that the smaller d the more it can be expected, that all solutions of the problems (Q$_k$) are the same and that by Theorem 5 this solution yields the solution of (Q).

References

1. U. Faigle, W. Kern and M. Streng, Note on the computational complexity of j-radii of polytopes in $\mathbb{R}^n$, *Mathematical Programming* 73, 1–5, 1996.
2. P. Gritzmann and V. Klee, Inner and outer j-radii of convex bodies in finite dimensional normed spaces, *Discrete Computational Geometry* 7, 255–280, 1992.
3. R. Hettich and P. Zencke, *Numerische Methoden der Approximation und Semi-infiniten Optimierung*, B.G. Teubner, 1982.
4. R. Horst and H. Tuy, *Global Optimization*, Springer Verlag, 1990.
5. C.G. Gibson, K. Wirthmüller, A.A. Du Plessis and E.J.N. Looijenga, Topological Stability of Smooth Mappings, *Lecture Notes in Mathematics*, Vol. 522, Springer Verlag, Berlin, 1976.
6. F. John, Extremum problems with inequalities as subsidiary conditions, *Studies and Essays: Courant Anniversary Volume*, Interscience publishers, New York, 187–204, 1948.
7. H. W. Jung, Ueber die kleinste kugel, die eine räumliche Figur einschliesst, *Journal für die reine und angewandte Mathematik* 123, 241–257, 1901.
8. P. Jonker, M. Streng and F. Twilt, Chebyshev Line Approximation: a Sensitivity Analysis, in *Parametric Optimization and Related Topics III*, J. Guddat, H. Th. Jongen, B. Kummer and F. Nočička (eds.) (Approximation and Optimization), Peter Lang Verlag, Frankfurt am Main, 303–322, 1993.
9. D. Klatte, On quantitative stability for non-linear minima, *Control and Cybernetics* 23 (1/2), 183–200, 1994.
10. H. Späth, Orthogonal least squares fitting with linear manifolds, *Numerische Mathematik* 86, 441–445, 1986.
11. G. Still and M. Streng, Optimality Conditions in Smooth Nonlinear Optimization, *Journal of Optimization Theory and Applications* 90 (3), 483–516, 1996.
12. M. Streng, Chebyshev Approximation of Finite Sets by Curves and Linear Manifolds, dissertation, University of Twente, 1993.
13. M. Streng and W. Wetterling, Chebyshev Approximation of a Point Set by a Straight Line, *Constructive Approximation* 10, 187–196, 1994.