Discrete linear Chebyshev approximation

Citation for published version (APA):

Document status and date:
Published: 01/01/1972

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.
Discrete Linear Chebyshev Approximation

by

L.S. de Jong

T.H.-Report 72-WSK-02

July 1972
Introduction

In this report an overdetermined system of linear equations

\[ Ax = b \]

is studied. It is assumed that \( A \in \mathcal{M}_{m,n}^r \), \( b \in \mathbb{R}^m \) and that \( m > r \). Generally, no \( x \in \mathbb{R}^n \) exists that solves the system. The alternative is to determine the set \( T \) of vectors \( x \in \mathbb{R}^n \) so that \( Ax - b \) in some norm is as small as possible.

Here the Chebyshev-norm is used and so we are interested in

\[ \rho := \min \max_{x \in \mathbb{R}^n} |(Ax-b)_i|, \]

and

\[ T := \{ x \in \mathbb{R}^n \mid \forall i, |(Ax-b)_i| \leq \rho \}. \]

We shall call \( T \) the set of minimax solutions of \( Ax = b \) and \( \rho \) the deviation. The sole condition that will be posed upon \( A \) is that the number of rows is greater than its rank.

It is well known that \( T \) is non-void and convex, and that it contains only one vector if \( A \) satisfies the Haar condition. Generally, \( T \) can be characterized as a polyhedron or a polytope, since it is the intersection of a number of half-spaces. This study enables us to predict the dimension of \( T \) and to characterize its extremal vectors.

Furthermore three algorithms will be discussed, capable of finding a minimax solution of \( Ax = b \).

The first one, due to Stiefel, is the classical algorithm that determines in case of Haar condition the minimax solution, while the second one, due to Descloux, is a variant, which determines also in case of non-Haar condition a minimax solution of \( Ax = b \). The third algorithm is another variant, which has certain advantages above the algorithm of Descloux. However, the finiteness of this algorithm is only conjectured.

Finally, the related problem:

find \( x \in \mathbb{R}^n \) so that \( \begin{cases} \begin{array}{c} A_1 x = b_1, \\ \| A_2 x - b_2 \|_\infty \text{ is minimal}, \end{array} \end{cases} \)

with \( A_1 \in \mathcal{M}_{r,n}^r \) and \( A_2 \in \mathcal{M}_{m,n}^n \), \( m > n \),

will be considered and an algorithm will be proposed.
1. In this section we shall give a number of definitions and lemmas, which will be useful later on. Although all these may be formulated for any Hilbert space, we shall, for obvious reasons, formulate them only for the special case of $\mathbb{R}^n$.

1.1. One of the notions that will be used is that of a convex hull.

1.1.1. Definition.
Let $V$ be a subset of $\mathbb{R}^n$.
CH($V$) is called the convex hull of $V$ if

$$\text{CH}(V) := \{x \in \mathbb{R}^n; x = \sum \alpha_i v_i, \sum \alpha_i = 1, \alpha_i \geq 0, v_i \in V\}.$$

1.1.2. Lemma.
Let $v_1, \ldots, v_m$ be vectors in $\mathbb{R}^n$.
The following two assertions are equivalent.

$$\emptyset \in \text{CH}(\{v_1, \ldots, v_m\}) \iff \exists u \in \mathbb{R}^n \{\forall i \in \{1, \ldots, m\}: u^Tv_i > 0\}.$$

Proof. See [1].

A notion that is very important when dealing with the minimax solutions of $Ax = b$, is that of a cadre. See [2,3].

1.1.3. Definition.
Let $V := \{v_1, \ldots, v_{k+1}\}$ be a subset of $\mathbb{R}^n$.
$V$ is called a cadre, if

i) the rank of the matrix made up by $v_1, \ldots, v_{k+1}$ equals $k$;

ii) the rank of the matrix made up by any $k$ distinct vectors of $V$ also equals $k$.

1.1.4. Lemma.
$V := \{v_1, \ldots, v_{k+1}\}$ is a cadre if and only if $v_1, \ldots, v_{k+1}$ are linearly dependent and any relation of the form

$$\sum_{i=1}^{k+1} \lambda_i v_i = 0, \sum_{i=1}^{k+1} |\lambda_i| > 0$$

implies that all $\lambda_i$ are unequal to zero.
Proof. First we shall prove the necessary part of the lemma.
Let $V$ be a cadre. Then, as a consequence of 1.1.3, the vectors $v_1, \ldots, v_{k+1}$ are linearly dependent.
Consider a relation
\[ \sum_{i=1}^{k+1} \lambda_i v_i = 0 \text{ with } \sum_{i=1}^{k+1} |\lambda_i| > 0. \]

If we assume that some $\lambda_i$, for instance $\lambda_1$, equals zero, then the rank of the matrix made up by $v_2, \ldots, v_{k+1}$ must be less than $k$. This is contrary to the definition of a cadre, so indeed all $\lambda_i$ are distinct from zero.

On the other hand, if $v_1, \ldots, v_{k+1}$ are linearly dependent and if any relation of the form
\[ \sum_{i=1}^{k+1} \lambda_i v_i = 0 , \sum_{i=1}^{k+1} |\lambda_i| > 0 \]
implies that all $\lambda_i$ are unequal to zero, then the rank of the matrix made up by $v_1, \ldots, v_{k+1}$ equals $k$ and the rank of the matrix made up by any $k$ distinct vectors of $V$ also equals $k$, since otherwise there would be a relation of the known form with some $\lambda_i$ equal to zero.

1.1.5. Corollary.
Let $V$ be a subset of $\mathbb{R}^n$ consisting of the vectors $v_1, \ldots, v_m$. If the rank $r$ of the corresponding matrix is less than $m$, then $V$ contains a subset that is a cadre.
If $r = m-1$, then $V$ contains exactly one cadre.

Proof. We know that $r < m$.
Let $v_1, \ldots, v_r$ be a linearly independent subset of $V$, then $v_1, \ldots, v_{r+1}$ are linearly dependent.
Consequently there are scalars $\lambda_i$ that are not all equal to zero so that
\[ \sum_{i=1}^{r+1} \lambda_i v_i = 0 . \]
The set
\[ W := \{ v_i \mid 1 \leq i \leq r+1, \lambda_i \neq 0 \} \]
is by 1.1.4 the only cadre contained in $\{v_1, \ldots, v_{r+1}\}$. 🟢
1.1.6. Lemma.

Let $V$ be a subset of $\mathbb{R}^n$ consisting of the vectors $v_1, \ldots, v_m$.
If $\emptyset$ belongs to the convex hull of $V$, then $V$ contains a cadre $K$ so that

$$\emptyset \in CH(K).$$

**Proof.**

$$\emptyset \in CH(V) \Rightarrow \emptyset = \sum_{i=1}^{m} a_i v_i, \quad \sum_{i=1}^{m} a_i = 1, \quad a_i \geq 0.$$ 

Let, for $i = 1, \ldots, m_0$, $a_i > 0$ and for $i = m_0+1, \ldots, m$, $a_i = 0$.

The rank of the matrix made up by $v_1, \ldots, v_{m_0}$ is less than or equal to $(m_0-1)$.

If the rank equals $(m_0-1)$, then $K := \{v_1, \ldots, v_{m_0}\}$, which for this case concludes the proof.

If the rank is less than $(m_0-1)$, then there are scalars $\beta_1, \ldots, \beta_{m_0-1}$ that are not all equal to zero so that

$$\sum_{i=1}^{m_0-1} \beta_i v_i = \emptyset.$$ 

Let $\gamma$ be an arbitrary scalar, then

$$\sum_{i=1}^{m_0} a_i v_i + \gamma \sum_{i=1}^{m_0-1} \beta_i v_i = \emptyset,$$

or

$$\sum_{i=1}^{m_0-1} (a_i + \gamma \beta_i) v_i + \alpha_m v_m = \emptyset.$$ 

Choose $\gamma$ so that at least one of the new coefficients is equal to zero and the others are non-negative. We obtain a proper subset that contains $\emptyset$ in its convex hull. With this subset we can go on in a similar way, the procedure terminating if the rank becomes one less than the number of vectors in it.

The procedure does terminate since every time a proper subset is formed.

The final subset contains at least one vector and, since the rank is one less than the number of vectors, the final subset is a cadre containing $\emptyset$ in its convex hull.

\[\Box\]
Remark.

This lemma is an extension of the theorem of Caratheodory, as formulated in [1]:
"If \{v_1,\ldots,v_m\} is a set of vectors in \(\mathbb{R}^n\) and \(m \geq n+1\), while \(\emptyset \in \text{CH}(\{v_1,\ldots,v_m\})\), then there is a subset of at most \((n+1)\)-vectors containing \(\emptyset\) in its convex hull."

1.2. One of our aims is to predict the dimension of \(T\), the set of minimax solutions, so we ought to give a definition that tells what dimensions subsets of \(\mathbb{R}^n\) have.
This will be done with the aid of hyperplanes.

1.2.1. Definition.

Let \(v_1,\ldots,v_d\) be a linearly independent set of vectors in \(\mathbb{R}^n\), let \(x_0 \in \mathbb{R}^n\), then
\[
V := \{x \mid x = x_0 + \sum_{i=1}^{d} \lambda_i v_i, \lambda_i \in \mathbb{R}\}
\]
is called a hyperplane of dimension \(d\).

This definition is equivalent with the following

1.2.2. Definition.

Let \(h_1,\ldots,h_{n-d}\) be a linearly independent set of vectors in \(\mathbb{R}^n\), let \(x_0 \in \mathbb{R}^n\), then
\[
V := \{x \mid (x-x_0)^T h_i, i = 1,\ldots,(n-d)\}
\]
is called a hyperplane of dimension \((n-d)\).

Remark.

According to the definition a subset consisting of only one vector is itself a hyperplane of dimension zero.

1.2.3. Definition.

Let \(V\) be a subset of \(\mathbb{R}^n\).
If \(V\) is contained in a hyperplane of dimension \(d\), but not in a hyperplane of lower dimension, then the dimension of \(V\) is \(d\).
We now have the tool to determine the dimension of subsets of $\mathbb{R}^n$: the dimension of a subset $V$ is equal to that of the "minimal" hyperplane containing $V$.

1.3. Already in the introduction we stated that the set of minimax solutions is a polyhedron or a polytope.

1.3.1. Definition.

Let $V_1, \ldots, V_p$ be half-spaces in $\mathbb{R}^n$.
(So $V_i := \{ x \mid x \in \mathbb{R}^n, x^T v_i \leq c_i; v_i \in \mathbb{R}^n, c_i \in \mathbb{R} \}$.)

If the intersection of $V_1, \ldots, V_p$ is nonempty, then, if $\bigcap_{i=1}^p V_i$ is bounded,

$\bigcap_{i=1}^p V_i$ is called a polyhedron, otherwise $\bigcap_{i=1}^p V_i$ is called a polytope.

An alternative way to define a polyhedron is by

1.3.2. Definition.

Let $V$ be a subset of $\mathbb{R}^n$.

If there are vectors $v_1, \ldots, v_m$ so that

$$V = \text{CH}(\{v_1, \ldots, v_m\}),$$

then $V$ is called a polyhedron.

See [4, 6, 7, 8, 9] for the proof that these definitions are equivalent.

1.3.3. Definition.

Let $V$ be a subset of $\mathbb{R}^n$ and let $v \in V$.

If

$$\forall v_1, v_2 (\neq v) \in V, \ v \notin \text{CH}(\{v_1, v_2\}),$$

then $v$ is called an extremal vector of $V$. 

1.3.4. Lemma.

Let $V$ be a non-void compact and convex set in $\mathbb{R}^n$.

(A set is convex if it contains its convex hull.)

Then

i) $V$ has extremal vectors,

ii) if there are finitely many extremal vectors, then $V$ is a polyhedron.

Proof. See [1, 4, 7, 8, 9].

A consequence of this lemma is that a polyhedron is identical with the convex hull if its extremal vectors.

Definition 1.3.3 will be used when characterizing the extremal vectors of $T$. 
2. In this section we shall be concerned with the system of linear equations

\[ Ax = b. \]

A will be a matrix having m rows and n columns and it will be assumed that the rank of A is less than m, so

\[ A \in \mathbb{R}^{m \times n}, \ b \in \mathbb{R}^m, \ r < m. \]

The rows of A will be denoted by

\[ a_1^T, \ldots, a_m^T. \]

2.1.

2.1.1. Definition.

i) \( x \in \mathbb{R}^n \) is called a minimax solution of \( Ax = b, \) if \( \| Ax - b \|_{\infty} \) is minimal.

ii) \( T(A, b) := \{ x \in \mathbb{R}^n \mid \| Ax - b \|_{\infty} \text{ is minimal} \}. \)

iii) \( \rho(A, b) := \min_{x \in \mathbb{R}^n} \| Ax - b \|_{\infty} \) is called the deviation.

iv) \( r(x) := Ax - b \) is called the residual vector to \( x \in \mathbb{R}^n. \)

2.1.2. Theorem.

\( T(A, b) \) is nonempty, convex and closed.

\( T(A, b) \) is bounded, if and only if \( r = n. \)

If \( r = n \) \( T(A, b) \) is a polyhedron, if \( r < n \) a polyhedral cylinder.

Proof. See [1].

Remark.

If \( r < n, \) we would expect \( T(A, b) \) to be a rather general polytope, since \( T(A, b) \) is the intersection of a number of half-spaces.

However, since along with

\[ \{ x \mid a_i^T x \leq b_i + \rho(A, b) \}, \]

also

\[ \{ x \mid a_i^T x \geq b_i - \rho(A, b) \} \]
occurs, $T(A, b)$ is the intersection of a finite number of "slabs" and it is not difficult to see that this results either in a polyhedron or in a cylinder.

In section 1 we produced evidence of:
"given a set of linearly dependent vectors in $\mathbb{R}^n$, then there is always a subset that is a cadre".

The rows of $A$ are linearly dependent elements of $\mathbb{R}^n$, so there is a number of rows that make up a cadre. The corresponding submatrix will be called a cadre matrix. Henceforth we shall be mostly dealing with subsystems of $Ax = b$ instead of the entire system. In order to be able to handle the subsystems more easily and also to formulate certain conditions which $A$ may satisfy more elegantly, we shall introduce a formalism (which can also partly be found in [10]).

2.1.3. Notation/definition.

Let $L$ be the set of the indices of the rows of $A$, so

$L := \{1, \ldots, m\}$.

Let $I$ be a subset of $L$ having the elements $i_1, \ldots, i_k$ so

$I := \{i_1, \ldots, i_k\} \subset L$.

Then

i) $\mu(I) := |I|$ (i.e. the number of elements in $I$).

ii) $A(I) := (a_{i_1}, \ldots, a_{i_k})^T$, so $A(I) \in \mathbb{R}^{\mu(I), n}$.

$b(I) := (b_{i_1}, \ldots, b_{i_k})^T$, so $b(I) \in \mathbb{R}^{b(I)}$.

iii) $\mathcal{I} := \{I \mid I \subset L\}$.

iv) $\mathcal{N}_k := \{I \mid I \in \mathcal{I} ; \mu(I) = k\}$.

v) $\mathcal{R}_k(A) := \{I \mid I \in \mathcal{N}_{k+1} ; \text{rank } A(I) = k\}$.

vi) $\mathcal{H}_k(A) := \{I \mid I \in \mathcal{R}_k(A) ; \forall I' \subset I : \text{rank } A(I') = \mu(I')\}$. 

□
Let us formulate the meaning of (v) en (vi):

Any element of $\mathcal{R}_k(A)$ determines a submatrix of $A$ having $(k+1)$ rows and rank $k$. These submatrices will be referred to as references.

Any element of $\mathcal{N}_k(A)$ determines a submatrix of $A$ that is a cadre matrix.

To illustrate the use of this formation some examples may be useful.

(i) "Let $(\bar{a}_1, \ldots, \bar{a}_k)^T$ be a submatrix of $A$ that is a cadre matrix".

Our new formulation for this clause is

"Let $I = \{i_1, \ldots, i_k\}$ and let $I \in \mathcal{N}_{k-1}(A)$".

(ii) "$A$ satisfies the Haar condition, if the matrix made up by any $(r+a)$ different rows of $A$ is a cadre matrix".

Our formulation becomes

"$A$ satisfies the Haar condition if $\mathcal{N}_r(A) = \mathcal{N}_{r+1}$".

This formalism will particularly in section 3 be valuable.

2.2. In the next pages first of all the case will be considered that $A$ is a cadre matrix. Then the finding of the manimax solutions of $A$ will appear to be an easy task. An explicit expression for the deviation exists and, moreover, we shall see that any $x \in T(A, b)$ has the same residual vector.

2.2.1. Theorem.

Let $A \in \mathcal{M}_{m,n}$, $L \in \mathcal{N}_r(A)$ (so $A$ is itself a cadre matrix). Then

(i) $r = m-1$,

(ii) there is a $\lambda \in \mathbb{R}^m$ so that $\lambda \neq 0$ and $\sum_{i \in L} \lambda_i a_i = 0$, $\lambda$ is uniquely determined up to multiplication by a scalar,

(iii) $\rho := \left( - \sum_{i \in L} \lambda_i b_i \right) \left( \sum_{i \in L} |\lambda_i| \right)^{-1}$, $\rho(A, b) = |\rho|$, $\rho(L)$.

(iv) $x \in T(A, b)$ if and only if

$$\forall i \in L, \ r_i(x) = \rho \text{ sign}(\lambda_i)$$

Proof.

(i) Obvious from definition 1.1.3.
(ii) The rows of \( A \) are linearly dependent and so there are scalars \( \lambda_i \) that are not all equal to zero so that
\[
\sum_{i \in L} \lambda_i a_i = 0.
\]
From lemma 1.1.4 we obtain that
for \( i \in L, \lambda_i \neq 0. \)

(iii) Consider the system of linear equations:
for \( i \in L, \ a_i^T x = b_i + \rho \text{sign}(\lambda_i). \)

Because
\[
\sum_{i \in L} \lambda_i b_i + \rho \sum_{i \in L} |\lambda_i| = 0,
\]
and
\[
\sum_{i \in L} \lambda_i a_i = \theta, \ rank \ A = r,
\]
it follows that the rank of the "extended" matrix is equal to that of \( A \) and so there is an \( x \in \mathbb{R}^n \) solving the system. This shows that \( \rho(A, b) \geq |\rho|. \)

On the other hand, for any \( x \in \mathbb{R}^n \) we have
\[
A^T \Gamma(x) = -A^T b = \rho \sum_{i \in L} |\lambda_i|.
\]

Hence,
\[
|\rho| = |A^T \Gamma(x)| \left( \sum_{i \in L} |\lambda_i| \right)^{-1} \leq \| \Gamma(x) \|_\infty
\]
and this proves (iii).

(iv) is proved by the observation that in (*) equality holds if and only if
\[
r_i(y) = \rho \text{sign}(\lambda_i).
\]

2.2.2. Theorem.
Let \( I \in \mathcal{J} \), then
\[
\rho(A(I), b(I)) \leq \rho(A, b).
\]

Proof. Obvious. □
Remark.

If \( \rho(A(I), b(I)) = \rho(A, b) \), then \( T(A(I), b(I)) \supset T(A, b) \); this conclusion is not necessarily true, however, if \( \rho(A(I), b(I)) < \rho(A, b) \).

In the foregoing we saw that for a cadre matrix the deviation can be found easily, and that the residual vector \( r(x) \) is constant for \( x \in T(A, b) \).

On the other hand, if \( A \) is not itself a cadre matrix, we know that \( A \) contains a submatrix that is.

The question, which can be asked is:

"is there such a cadre matrix that the deviation of the corresponding subsystem is equal to that of the whole system?".

We shall prove that the answer is positive. Let us call such a cadre matrix a maximal cadre matrix. Solving the minimax problem for the subsystem corresponding with a maximal cadre matrix then supplies

(i) the deviation of the whole system,
(ii) a set of minimax solution that contains the set of minimax solutions of the whole system,
(iii) the knowledge that for any minimax solution \( x \) of the whole system the components of \( r(x) \), corresponding with the subsystem to the maximal cadre matrix, are the same.

2.2.3. Definition.

Let \( K \in \mathcal{J} \) and \( \mu(K) = k+1 \).

\( A(K) \) is called a maximal cadre matrix with respect to \( b \), if

(i) \( K \in \mathcal{A}(A) \),

(ii) \( \rho(A(K), b(K)) = \rho(A, b) \).

2.2.4. Theorem.

There is a \( K \in \mathcal{J} \) so that \( A(K) \) is a maximal cadre matrix with respect to \( b \).

Proof. Let \( x \in T(A, b) \) and define \( I \in \mathcal{J} \) and \( \sigma_i \) by

\[
I := \{ i \mid |r_i(x)| = \rho(A, b) \}
\]

\[
\sigma_i := \text{sign}(r_i(x)), \ i \in I.
\]

Since \( T(A, b) \) is closed, \( I \) is non-void.
There is no \( u \in \mathbb{R}^n \) so that
\[
\text{for } i \in I, \sigma_i a_i u > 0,
\]
for, if there was, then for \( \varepsilon \) small enough and positive we would have
\[
\| x - y \|_{\infty} < \rho(A, b),
\]
and this is contrary to \( x \in T(A, b) \).

The consequence (see lemma 1.1.2) is that
\[
\theta \in \text{CH}(\{\sigma_i a_i | i \in I\}).
\]

According to 1.1.6 there is a subset \( K \) of \( I \) so that
\[
\begin{align*}
(i) & \quad \theta \in \text{CH}(\{\sigma_i a_i | i \in K\}) \\
(ii) & \quad K \in \mathcal{H}_k(A).
\end{align*}
\]

Consequently there are scalars \( \mu_i \) that are not all equal to zero so that
\[
\theta = \sum_{i \in K} \mu_i \sigma_i a_i,
\]
and now we obtain
\[
\rho(A(K), b(K)) = \left| - \sum_{i \in K} \mu_i \sigma_i b_i \right| \left( \sum_{i \in K} |\mu_i| \right)^{-1} =
\]
\[
- \sum_{i \in K} \mu_i \rho(A, b) \left( \sum_{i \in K} |\mu_i| \right)^{-1} = \rho(A, b).
\]

This last proves the theorem. \( \square \)

2.2.5. Theorem.

If \( r = m-1 \), then there is only one \( K \) so that \( A(K) \) is a maximal cadre matrix with respect to \( b \).

This \( K \) is the same for any \( b \in \mathbb{R}^m \).

Proof. From lemma 1.1.5 it follows that \( A \) contains one cadre matrix if \( r = m-1 \).

Applying theorem 2.2.4, the required result follows. \( \square \)
2.3. In this part attention will be paid to the set of minimax solutions $T(A, b)$ and, again, first of all the case will be considered, that $A$ is a cadre matrix.

2.3.1. Theorem.
Let $L \in \mathcal{M}_{r+1}(A)$, then $T(A, b)$ is a subspace of dimension $(n-r)$.

Proof. Let $x \in T(A, b)$, by theorem 2.2.1 this is equivalent to:
$$\sum_{i \in L} a_i T x - b_i = \rho \text{sign}(\lambda_i).$$

The dimension of the set of solutions of this system is $(n-r)$.

2.3.2. Theorem.
Let $A$ have only one maximal cadre matrix $A(K)$ with respect to $b$ and let $\mu(K) = k+1$. Then the dimension of $T(A, b)$ is equal to $(n-k)$.

Proof. The case $r = n$ is considered in theorem 2.3.1, so let us assume that $r < n$.

For the sake of brevity we shall denote $T(A(I), b(I))$ and $\rho(A(I), b(I))$, for any $I \in \mathcal{I}$, as $T(I)$ and $\rho(I)$. From theorem 2.3.1 it follows that
$$\dim T(K) = (n-k).$$

We shall show that the supposition
$$\dim T(L) < (n-k)$$

contradicts the assumption that $A$ has only one maximal cadre matrix. The proof will be constructed in a number of steps.

(i) From (*) it follows at once that there is an $I$ so that

(a) $K \subset I \subset L$,
(b) $\dim T(I) < (n-k)$,
(c) for any $I'$ so that $K \subset I' \subset I$ holds: $\dim T(I') = (n-k)$.

(ii) The next step is to prove:

For all $i \in I$ there are scalars $\sigma_i$ with $\sigma_i = \pm 1$ so that
$$\forall x \in T(I), \forall i \in I, r_i(x) = \rho(I)\sigma_i.$$.
Proof. For \( i \in K \) the assertion is trivial. Let \( p \in I K \) and \( I' := I \backslash \{p\} \).

As dimension \( T(I) < (n-k) \), there is, according to 1.2.3, a hyperplane \( V \) of dimension \( (n-k-1) \) so that \( V \supset T(I) \). Since \( \dim T(I') = (n-k) \), we obtain that \( T(I') \backslash V \) is non-empty. Let \( \gamma \in T(I') \backslash V \) and let \( x \in T(I) \), then

\[
\begin{align*}
\text{for } i \in I, \quad |r_i(x)| &\leq \rho(I) \quad \text{(since } x \in T(L)), \\
\text{for } i \in I', \quad |r_i(\gamma)| &\leq \rho(I') = \rho(I) \quad \text{(since } \gamma \in T(I')), \\
|r_p(\gamma)| &> \rho(I) \quad \text{(since } \gamma \notin T(I')).
\end{align*}
\]

Now, if \( |r_p(\gamma)| < \rho(I) \), then there is a scalar \( \alpha \in (0,1) \) so that \( \alpha x + (1-\alpha)\gamma \in T(I) \) and this means that, since \( T(I) \) is contained in the hyperplane \( V \), \( \gamma \in V \), but this is not true.

Therefore
\[
|r_p(x)| = \rho(I).
\]

Because \( x \in T(I) \) and \( p \in I \backslash K \) were chosen arbitrarily, (ii) has been proved. (The \( \sigma_i \) do not depend on \( x \) since otherwise \( x^* \in T(I) \) would exist, not satisfying (ii).)

(iii) There is no \( u \in \mathbb{R}^n \) so that:

\[
\begin{align*}
\text{for } i \in K, \quad a_i^Tu &= 0 \\
\text{for } i \in I \backslash K, \quad \sigma_i a_i^Tu &> 0.
\end{align*}
\]

Proof. If \( u \) did exist there would be an \( x \in T(I) \) not satisfying (ii).

(iv) Let for \( i \in I \) \( a_i \) be written as:

\[
a_i = a_i^1 + a_i^\|, \quad \text{where}
\]

\[
a_i^\| \in S := \text{span of } \{a_i \mid i \in K\} \quad \text{and}
\]

\[
a_i^1 \in S^1,
\]

then

\[
\theta \in CH(\{\sigma_i a_i^1 \mid i \in T|K\}).
\]

Proof. Suppose that

\[
\theta \notin CH(\{\sigma_i a_i^1 \mid i \in I|K\}),
\]

then there is a \( u \in \mathbb{R}^n \) so that
for $i \in I \setminus K$, $(\sigma_i \bar{a}_i \mathbf{u})^T > 0$  \hspace{1cm} \text{(lemma 1.1.2)}

or

for $i \in I \setminus K$, $(\sigma_i \bar{a}_i \mathbf{u})^T > 0$

and

for $i \in K$, $\bar{a}_i \mathbf{u}^T = 0$,

and this contradicts (iii).

The consequence of (iv) is that there are scalars $a_i \geq 0$ so that

\[ \theta = \sum_{i \in I \setminus K} \sigma_i a_i \bar{a}_i^T + \sum_{i \in I \setminus K} a_i = 1. \]

Let $\bar{a}^* := \sum_{i \in I \setminus K} \sigma_i a_i \bar{a}_i^T$, then $\bar{a}^* = \theta$ and thus $\bar{a}^* \in S$.

(v) There is an $\ell \in K$ so that

\[ \emptyset \in CH(\{\sigma_i a_i \mid i \in I \setminus \{\ell}\}). \]  \hspace{1cm} \text{(**)}

**Proof.** $A(K)$ is a cadre matrix. According to theorem 2.2.1 there is a $\lambda \neq \emptyset$, with all $\lambda_i \neq 0$, so that

\[ \sum_{i \in K} |\lambda_i| \sigma_i a_i = \emptyset. \]

$\bar{a}^* \in S$, so there are scalars $b_i$ so that

\[ \sum_{i \in K} -b_i \sigma_i a_i = \bar{a}^*. \]

Hence we have for any scalar $\gamma$

\[ \sum_{i \in I \setminus K} \sigma_i a_i \bar{a}_i^T + \sum_{i \in K} (\beta_i + \gamma |\lambda_i|) \sigma_i a_i = \emptyset. \]

Choose $\gamma$ so that for $i \in K$, $\beta_i + \gamma |\lambda_i| \geq 0$ and that for at least one $\ell \in K$, $\beta_\ell + \gamma |\lambda_\ell| = b$ (\(\gamma := \max(-\frac{\beta_i}{|\lambda_i|})\)).

The consequence is

\[ \emptyset \in CH(\{\sigma_i a_i \mid i \in I \setminus \{\ell\}, \ell \in K\}). \]  \hspace{1cm} \Box

In a similar way as in the proof of theorem 2.2.1 it now follows from (**) that $A$ contains a maximal cadre matrix different from $A(K)$ (since this new cadre does
not contain $\ell \in K$, and therefore the supposition
\[ \dim T(L) < (n-k) \]
is incorrect.

2.3.3. Theorem.
Let $A$ have the maximal cadre matrices $A(K_1), \ldots, A(K_s)$ with respect to $b$.

If $K := \bigcup K_i$ and $\ell := \dim T(A(K_i), b(K_i))$ then
\[ \dim T(A, b) = \ell. \]

Proof. The theorem is the extension of 2.3.2 for the case that $A$ has more than one maximal cadre matrix.
The proof follows the same lines except that in part (v) some extra care is needed, since there are more cadre matrices and consequently the $u$-vector has more than one degree of freedom.

2.4. From theorem 2.1.2 we know that $T(A, b)$ is bounded and, more specifically, a polyhedron if and only if the rank of $A$ equals $n$.
It should be observed that the difference between the case $r = n$ and the case $r < n$ is not essential. If $A \in M_{m,n}$ with $r < n$, then one can find $L \in M_{r,n}$ and $U \in M_{n,n}$ so that $A = L \times U$. $L$ has full rank and thus the minimax solutions of $Ly = b$ constitute a polyhedron.
But any minimax solution of $Ly = b$ introduces via $Ux = y$ an $(n-r)$-dimensional hyperplane of minimax solutions of $Ax = b$.
Conversely, with any minimax solution of $Ax = b$ corresponds a minimax solution of $Ly = b$ via $y = Ux$. We can also formulate this as follows:
If $r < n$ then $T(A, b)$ is the direct sum of a polyhedron and the kernel of $A$.
For the rest of this section we only consider the case $r = n$.

2.4.1. Definition.
$x$ is called an extremal minimax solution of $Ax = b$, if $x$ is an extremal vector of $T(A, b)$.

2.4.2. Theorem.
Let $x_0 \in T(A, b)$ and $I := \{i \mid r_i(x_0) = \rho(A, b)\}$, then $x_0$ is an extremal minimax solution if and only if
\[ \text{rank } A(I) = n. \]
Remark.

If rank $A(I) = n$, then $\mu(I) \geq n+1$, since $A(I)$ must contain a maximal cadre matrix.

Proof. A solution of the system

$$i \in I, \ a_i^T x - b_i = \rho(A,b) \sigma_i$$

is determined modulo the kernel of $A(I)$.

If rank $A(I) < n$, then the kernel is not trivial and so there is a $y \neq \emptyset$ so that

$$A(I)y = \emptyset.$$

Now, if $|I|$ is small enough, then $x_1 := x_0 + \delta y$ is a minimax solution of $Ax = b$ and this contradicts the assumption that $x_0$ is an extremal vector of $T(A,b)$, $(x_0 = \frac{1}{2}(x_0 + \delta y) + \frac{1}{2}(x_0 - \delta y))$ and so indeed

$$\text{rank } A(I) = n.$$

On the other hand, if rank $A(I) = n$, we must show that $x_0$ is an extremal minimax solution.

Let $x_1 \in T(A,b)$ and $x_2 \in T(A,b)$ so that

$$x_0 = \frac{1}{2} x_1 + \frac{1}{2} x_2,$$

then

$$\text{for } i \in I, \ \frac{1}{2} r_i(x_1) + \frac{1}{2} r_i(x_2) = \sigma_i \rho(A,b)$$

and so

$$\text{for } i \in I, \ r_i(x_1) = r_i(x_2) = r_i(x_0) = \sigma_i \rho(A,b).$$

Rank $A(I) = n$, therefore the solution of the system

$$i \in I, \ a_i^T x - b_i = \sigma_i \rho(A,b)$$

is uniquely determined, so $x_0 = x_1 = x_2$ and this implies that $x_0$ is an extremal vector of $T(A,b)$.

Remark.

By a combinatorial argument, it follows from 2.4.6 that $T(A,b)$ has only finitely many extremal vectors. Since $T(A,b)$ is compact we now obtain in another way that $T(A,b)$ is a polyhedron.
2.4.3. Theorem.

Let \( A \in \mathcal{M}_{n+1,n} \). Then there exists a cube \( C \) in \( \mathbb{R}^n \) so that there is a 1-1 linear mapping from \( C \) to \( T(A,b) \).

Proof. Let \( A(K) \) be the maximal cadre matrix of \( A \). Since \( A \in \mathcal{M}_{n+1,n} \), \( K \) is indeed a cadre (the only one \( A \) has).

Let \( \ell \in K \), then

\[
T(A,b) = \{ y \mid y \in \mathbb{R}^n, i \in K \setminus \ell, a_i^T x - b_i = \sigma_i \rho(A,b), \ |\sigma_i| = 1 \}
\]

Let \( C := \{ y \mid y \in \mathbb{R}^n, i \in K \setminus \ell, y_i - b_i = \sigma_i \rho(A,b), \ |\sigma_i| = 1 \}
\]

is a cube in \( \mathbb{R}^n \) and the mapping from \( C \) to \( T(A,b) \) defined by

\[
x = A(I \setminus \ell) y^{-1}
\]

is clearly 1-1 and linear. \( \square \)

The consequence of the theorem is that in the case of \( A \in \mathcal{M}_{n+1,n} \), and if \( \mu(K) = k \), \( T(A,b) \) is a \((n-k)\)-dimensional parallelepiped with \( 2^{n-k} \) extremal vectors.

If the number of rows of \( A \) is greater than \((n+1)\), then the only thing that can be said about the number of extremal vectors is that it is at least equal to \((n-k+1)\) and that it may be greater than \( 2^{n-k} \), a trivial upperbound being \( \binom{n-k-1}{n-k} 2^{n-k} \).

These bounds are illustrated by figure 1.

The plane of drawing corresponds with \( T(A(K),b(K)) \) and \( T(A,b) \) is the intersection of a number of "slabs".
\[ m = n+2, \text{ 3 extremal vectors} \quad \text{and} \quad m = n+2, \text{ 6 extremal vectors} \]

\[ n = 3 \quad , \quad k = 1 \]

\text{figure 1.}
3. In this section we shall be concerned with the computation of a minimax solution for $Ax = b$ with $A \in \mathbb{R}_{m,n}^n$ and $m > n$. If $m = n + 1$, it is obvious how this should be done (section 2). However, if $m > n + 1$, it is also clear that the task is a complicated one. What we would like is to determine a reference (by this we mean an $A(I), I \in \mathcal{R}_n(A)$), for which the deviation is equal to the deviation of the whole system, and to determine a minimax solution for this reference, which is also a minimax solution for the whole system. Thus, what we need is an algorithm that in a finite number of steps finds such a (maximal) reference. The classical algorithm of Stiefel does so for the case that $A$ satisfies the Haar condition but may fail otherwise.

We intend to show why the classical algorithm need not be finite if $A$ does not satisfy the Haar condition, and shall discuss a variant invented by Descloux, and prove that it supplies a minimax solution of $Ax = b$ after a finite number of steps. Furthermore we shall propose a variant of the algorithm of Descloux, having certain advantages. However, the finiteness of this variant is only conjectured.

3.1. Before discussing the various algorithms let us introduce a new tool and prove some theorems.

3.1.1. Definition.

Let $A \in \mathbb{R}_{m,n}^n$, $b \in \mathbb{R}^m$ and $K \in \mathcal{R}_k(A)$.

Let $\lambda \neq 0$ and $\sum_{i \in K} \lambda_i a_i = 0$, then

$$S(A(K),b(K)) := \{x | x \in \mathbb{R}^n; \exists \sigma \forall i \in k, \text{sign}(r_i(x)\lambda_i) = \sigma\}.$$  

In this definition it is left possible that $\sigma$ depends on the choice of $x$. However, this is not so and it is possible to find an expression for $\sigma$ in terms of the $\lambda_i$ and the $b_i$.

Obviously, we have

$$S(A(K),b(K)) \supseteq T(A(K),b(K)),$$

and this means that $S(A(K),b(K))$ is non-void.

For the sake of brevity, let $S(K)$ denote $S(A(K),b(K))$. 
3.1.2. Theorem.
For any \( x \in S(K) \)
\[
\sigma = \text{sign}(\sum_{i \in K} \lambda_i b_i).
\]

Proof. We know that for any \( x \) in \( \mathbb{R}^n \)
\[
- \sum_{i \in K} \lambda_i b_i = \sum_{i \in K} \lambda_i r_i(x).
\]
If \( x \in S(K) \), then we obtain
\[
\text{sign}(\sum_{i \in K} \lambda_i b_i) = \text{sign}(\sum_{i \in K} \lambda_i r_i(x)) =: \sigma.
\]

3.1.3. Corollary.
If \( x_1 \) and \( x_2 \in S(K) \), then
for \( i \in K \), \( \text{sign}(r_i(x_1)) = \text{sign}(r_i(x_2)) \).

3.1.4. Theorem.
Let
\[
\rho := (\sum_{i \in K} \lambda_i b_i) \left( \sum_{i \in K} |\lambda_i| \right)^{-1},
\]
then
\[
\forall x \in S(K), \min_{i \in K} |r_i(x)| \leq |\rho| \leq \max_{i \in K} |r_i(x)|,
\]
with either twice equality or twice inequality.

Proof. Since for any \( x \)
\[
- \sum_{i \in K} \lambda_i b_i = \sum_{i \in K} \lambda_i r_i(x)
\]
it follows that
\[
\rho = \left( \sum_{i \in K} \lambda_i r_i(x) \right) \left( \sum_{i \in K} |\lambda_i| \right)^{-1}.
\]
If \( x \in S(K) \), then
\[
|\rho| = \left( \sum_{i \in K} |\lambda_i| |r_i(x)| \right) \left( \sum_{i \in K} |\lambda_i| \right)^{-1},
\]
and from this the proof of the proposed theorem follows readily.
If any \((n+1)\) distinct rows of \(A\) make up a cadre matrix, then \(A\) satisfies the Haar condition. In section 2 we remarked that this statement could also be formulated as

\[
\mathcal{H}_n(A) = \mathcal{N}_{n+1}.
\]

We intend to show that this is true if any reference \(A(I)\) satisfies the Haar condition or, equivalently, is a cadre matrix.

3.1.5. Theorem.

\[
\mathcal{H}_n(A) = \mathcal{N}_{n+1} \iff \mathcal{R}_n(A) = \mathcal{H}_n(A).
\]

Proof. From \(\mathcal{N}_{n+1} = \mathcal{H}_n(A)\) it follows that \(\mathcal{R}_n(A) = \mathcal{H}_n(A)\), since

\[
\mathcal{N}_{n+1} \supset \mathcal{R}_n(A) \supset \mathcal{H}_n(A).
\]

Let \(\mathcal{R}_n(A) = \mathcal{H}_n(A)\) and let us suppose that \(\mathcal{H}_n(A)\) is a proper subset of \(\mathcal{N}_{n+1}\).

Then there are \((n+1)\) rows of \(A\) not making up a cadre matrix. These rows therefore contain a cadre \(K \notin \mathcal{H}_n(A)\). Since \(A\) has rank \(n\), \(K\) can be supplemented to become a reference \(I'\) not contained in \(\mathcal{H}_n(A)\).

This contradicts \(\mathcal{R}_n(A) = \mathcal{H}_n(A)\) and so indeed

\[
\mathcal{N}_{n+1} = \mathcal{H}_n(A).
\]

Remark.

The following does not hold:

\[
\mathcal{N}_{n+1} = \mathcal{R}_n(A) \iff \mathcal{N}_{n+1} = \mathcal{H}_n(A).
\]

3.2. We shall now formulate and prove the theorem that justifies the classical algorithm when applied in the case of Haar condition. Afterwards a similar theorem will be proved for non-Haar condition and it will be shown that the classical algorithm for this case need not be finite.

3.2.1. Theorem.

Let \(A \in \mathcal{H}_{n+2,n}\), \(b \in \mathbb{R}^{n+2}\) and let \(\mathcal{H}_n(A) = \mathcal{R}_n(A)\).

Let \(I_1 \in \mathcal{R}_n(A)\) and \(\rho(A(I_1), b(I_1)) > 0\).
Since $I_1 \in \mathcal{H}_n(A)$, $T(A(I_1), b(I_1))$ contains only one vector $x_1$.

Let $\{l\} := L|I_1$ and $|r_k(x_1)| > \rho(A(I_1), b(I_1))$.

Then there is exactly one $I_2 \in \mathcal{R}_n(A)$ so that

$$x_1 \in S(I_2).$$

For this $I_2$ holds

$$\rho(A(I_2), b(I_2)) > \rho(A(I_1), b(I_1)).$$

**Proof.** Let $\lambda \neq \emptyset$ so that $\sum_{i \in I_1} \lambda_i a_i = \emptyset$ and define, $\lambda \subseteq 0$.

Without loss of generality we may assume that

for $i \in I_1$, $r_i(x) = \rho(A(I_1), b(I_1)) \text{ sign } (\lambda_i)$, \hspace{1cm} \(\star\)

for if this is not so, then we may take $\lambda := -\lambda$.

Let $v$ be such that $\sum_{i \in I_1} v_i a_i = -\lambda$ and define $v \subseteq -1$.

Let for $s \in I_1$

$$I_2(s) := (I_1 \{s\}) \cup \{l\}.$$ 

Since $I_1 \in \mathcal{H}_n(A) = \mathcal{R}_n(A)$,

$$I_2(s) \in \mathcal{H}_n(A).$$

Let $u \neq \emptyset$ and $\sum_{i \in I_2(s)} u_i a_i = 0$ then $u = a \lambda - v$, with $a = \frac{v^s}{\lambda}$.

From 3.1.2 we have

$$x_1 \in S(I_2) \text{ if and only if}$$

for $i \in I_2(s) \\{l\}$, sign $(\lambda_i \frac{v^s}{\lambda} - \frac{v_i}{\lambda_i}) r_i(x_1)) = \text{ sign } (r_{\lambda}(x_1))$

or, using $(\star)$ and the fact that $I_2(s) \\{l\} = I_1 \\{l\}$,

for $i \in I_1 \\{l\}$, sign $(r_{\lambda}(x_1)\frac{v^s}{\lambda} - \frac{v_i}{\lambda_i})) = 1$. 


Sufficient and necessary for this is that

\[ r_\lambda(x_1) \frac{v_s}{\lambda_s} = \max_{i \in I_1} \left( r_\lambda(x_1) \frac{v_i}{\lambda_i} \right). \tag{**} \]

(**) determines \( s \in I_1 \) uniquely, for if \( \frac{v_s}{\lambda_s} \) were equal to \( \frac{v_t}{\lambda_t} \) and \( s \neq t \), then \( I_2(s) \in R_n(A) \), but not in \( R_n(A) \).

Since

\[
\begin{cases}
\min_{i \in I_2(s)} |r_i(x_1)| = \rho(A(I_1),b(I_1)) \\
\max_{i \in I_2(s)} |r_i(x_1)| = |r_k(x_1)| > \rho(A(I_1),b(I_1)),
\end{cases}
\]

it follows from theorem 3.1.4 that

\[ \rho(A(I_2),b(I_2)) > \rho(A(I_1),b(I_1)). \]

Remark.

If was given that \( I_1 \in R_n(A) \) instead of \( R_n(A) = R_n(A) \), then the choice of \( s \) will be in general not unique. However, with the same choice of \( s \), one would still obtain \( \rho(A(I_2),b(I_2)) > \rho(A(I_1),b(I_1)) \).

Remark.

We assumed that \( \rho(A(I_1),b(I_1)) > 0 \).
If \( \rho(A(I_1),b(I_1)) = 0 \), then the choice of \( I_2(s) \) is not unique, but if we choose \( s \) as in (**), then it follows that \( \rho(A(I_2(s)),b(I_2(s))) > \rho(A(I_1),b(I_1)) \).

3.2.2. The classical algorithm.

The task is the calculation of the minimax solution of \( Ax = b \) with \( A \in \mathbb{M}_{m,n} \), \( b \in \mathbb{R}^m \), \( m > n \) and \( R_n(A) = R_n(A) \).

The algorithm finds a maximal reference after a number of iterations and, since in this case the minimax solution of the maximal reference is unique, the unique minimax solution for the entire system.

The difference between consecutive iterations is that the corresponding references are equal except for one row. Let us suppose that in the \( j \)-th iteration the reference \( A(I_j) \) occurs and let the corresponding minimax solution be \( x_j \).
There are two possibilities. Either the maximal residual $\|r(x_j)\|_\infty$ is less than or equal to the current deviation, or the maximal residual is greater than the deviation. In the first case $x_j$ is clearly a minimax solution for the entire system. In the second case, suppose an $s \in I_j$ is chosen so that $I_{j+1} := (I_j \cup \{s\}) \cup \{t\}$ is the next reference.

If $x_j \notin S(I_{j+1})$, then it is not clear whether the deviation increases or decreases. In fact, both the one thing or the other is possible. However, since $A(I_j \cup \{s\})$ satisfies the Haar condition, $s$ can, according to theorem 3.2.1, be so chosen that $x_j \in S(I_{j+1})$. This choice of $s$ guarantees that $\rho(A(I_{j+1}), b(I_{j+1})) > \rho(A(I_j), b(I_j))$ and this explains why indeed in the classical algorithm $I_{j+1}$ is determined so that $x_j \in S(I_{j+1})$.

Since in any iteration the deviation strictly increases with this exchange criterion, the number of iterations is finite.

**Remark.**

Necessary for the classical algorithm to be finite is that the residual of the equation to enter the reference should be greater than the current deviation, not that it should be the maximal residual!

**Remark.**

Necessary for the application of theorem 3.2.1 is that at any iteration $A(I_j)$ satisfies the Haar condition, which is a weaker condition than $\mathcal{R}_n(A) = \mathcal{R}_n(A)$. For the algorithm of Descloux use will be made of this weaker condition when showing the finiteness.

### 3.2.3. A further algorithmic elaboration.

Let $A \in \mathbb{H}_{m,n}$, $m > n$ and let $\mathcal{R}_n(A) = \mathcal{R}_n(A)$, then

(0) let $j = 0$,

(1) choose $I \in \mathcal{R}_n(A)$,

(2) determine $\lambda \neq \emptyset$ so that

$$\sum_{i \in I} \lambda_i \neq \emptyset$$
and compute
\[ \rho(I) := \left(-\sum_{i \in I} \lambda_i b_i \right) \left(\sum_{i \in I} |\lambda_i|\right)^{-1}, \]

after which, if necessary, \( \lambda \) is adapted so that \( \rho(I) \geq 0 \),

(3) determine

for \( i \in I \),
\[ r_i := \rho(I) \text{ sign } (\lambda_i), \]

(4) compute the solution \( x \) of

\[ A(I) x = b(I) + r(I), \]

(5) determine \( \ell \in L/I \) so that \( |r_{\ell}(x)| \) is maximal, if \( |r_{\ell}(x_1)| \leq \rho(I) \)
then \( x \) is a minimax solution of \( Ax = b \), otherwise,

(6) compute \( \nu \) so that
\[ a_{\ell} = \sum_{i \in I} \nu_i a_i, \]

(7) determine \( s \in I \) so that
\[ r_{\ell}(x) \frac{\nu_s}{s} = \max_{i \in I} \left( r_{\ell}(x) \frac{\nu_i}{\lambda_i} \right), \]

(8) \( I := (I \{s\}) \cup \{\ell\}, j := j+1 \) and return to (2).

3.3. The case that \( \mathcal{R}_n(A) \) may not be identical with \( \mathcal{A}_n(A) \) is considered.

3.3.1. Theorem.

Let \( A \in \mathcal{A}_{n+2,n}^n, b \in \mathbb{R}^{n+2}. \)

Let \( I_1 \in \mathcal{R}_n(A) \) and \( K_1 \in \mathcal{R}_{K}(A) \) so that \( I_1 \supset K_1. \) Let \( \rho(A(I_1),b(I_1)) > 0. \)

Let \( x_1 \) be an extremal vector of \( T(A(I_1),b(I_1)). \)

Let \( \{\ell\} := L/I_1 \) and \( |r_{\ell}(x_1)| > \rho(A(I_1),b(I_1)). \)

Then there is a (generally not uniquely determined) \( I_2 \in \mathcal{R}_n(A), \)
distinct from \( I_1, \) so that, if its cadre is \( K_2, \)
\[ x_1 \in S(K_2). \]

For such an \( I_2 \) holds
\[ \rho(A(I_2),b(I_2)) \geq \rho(A(I_1),b(I_1)). \]
Proof. Let $\lambda \neq 0$ so that $\sum_{i \in K_1} \lambda_i a_i = 0$ and define for $i \in L|K_1$, $\lambda_i := 0$.

Without loss of generality we may assume

$$i \in K_1, r_i(x) = \rho(A(I_1), b(I_1)) \operatorname{sign}(\lambda_i), \quad (*)$$

for if this is not so, then we may take $\lambda := -\lambda$.

Let $\nu$ by such that $\sum_{i \in I_1} \nu_i a_i = \bar{a}_x$ and define $\nu_x := -1$.

We distinguish two cases.

1st. case: $\forall i \in I_1 | K_1, \nu_i r_i(x) \leq 0$ \hfill (**)

Choose $s \in K_1$ so that

$$r_s(x) \frac{\nu_s}{\lambda_s} = \max_{i \in K_1} (r_i(x) \frac{\nu_i}{\lambda_i})$$

(compare 3.2.1).

Consider

$$I_2 := (I_1 | \{x\}) \cup \{\ell\}.$$

It is obvious that $I_2$ is a reference and that $s$ need not be unique.

Let $K_2$ be the cadre of $I_2$, then $x \in K_2$.

Let $\mu \neq 0$ so that $\sum_{i \in K_1} \mu_i a_i = 0$, then $\mu = \alpha \lambda - \beta \nu$ with $\alpha \lambda_s = \beta \nu_s$ and $$(\alpha, \beta) \neq (0, 0).$$

Since $\lambda_s \neq 0$ we may choose $\beta = 1$ and hence have

$$\begin{cases} 
  \text{for } i \in K_2 \cap K_1, \mu_i = \lambda_i \left(\frac{\nu_s}{\lambda_s} - \frac{\nu_i}{\lambda_i}\right), \\
  \text{for } i \in K_2 \setminus K_1, \mu_i = -\nu_i.
\end{cases}$$

Using $(*)$ and $(**)$ we obtain

$$\text{for } i \in K_2, \operatorname{sign}(\mu_i r_i(x)) = \operatorname{sign}(r_i(x)),$$

which implies that $x \in S(K_2)$.

Since

$$\begin{cases} 
  \min_{i \in K_2} |r_i(x)| = \rho(A(I_1), b(I_1)), \\
  \max_{i \in K_2} |r_i(x)| = |r_\ell(x)| > \rho(A(I_1), b(I_1)),
\end{cases}$$

we have

$$\rho(A(I_1), b(I_1)) = \rho(A(I_1), b(I_1)).$$
it follows from theorem 3.1.4 that
\[ \rho(A(I_2), b(I_2)) > \rho(A(I_1), b(I_1)). \]

2nd. case: \( \exists t \in I_1 \mid \forall r \in S(x_1)x_2(x_1) > 0. \)

Consider
\[ I_2 := (I_1 \setminus \{t\}) \cup \{t\}. \]

Since \( v_t \neq 0, I_2 \) must be a reference.

\( I_1 \) and \( I_2 \) have the same cadre \( K_1 \), so \( x_1 \in S(K_2) = S(K_1) \), and, moreover,
\[ \rho(A(I_2), b(I_2)) = \rho(A(I_1), b(I_1)). \]

3.2.3. Corollary.

\( T(A(I_2), b(I_2)) \) has an extremal vector \( x_2 \) so that

for \( i \in I_2 \), \( \text{sign } (r_i(x_2)) = \text{sign } (r_i(x_1)). \)

Proof. Let \( x_2 \in T(A(K_2), b(K_2)) \), then \( x_2 \in S(K_2) \) or, equivalently,

for \( i \in K_2 \), \( \text{sign } (r_i(x_2)) = \text{sign } (r_i(x_1)). \)

For the equations not corresponding with the cadre matrix of \( A(I_2) \), the signs of the components of the residual vector are free.

Therefore choose

for \( i \in I_2 \setminus K_2 \), \( \text{sign } (r_i(x_2)) = \text{sign } (r_i(x_1)), \)

and this completes the proof.

3.3.3. Corollary.

There is an \( I_2 \), having a cadre \( K_2 \), satisfying
\[
\begin{align*}
\left\{ \begin{array}{l}
x_1 \in S(K_2), \\
\rho(A(I_2), b(I_2)) > \rho(A(I_1), b(I_1))
\end{array} \right. \\
\end{align*}
\tag{*}
\]

if and only if
\[
\forall i \in I_1 \mid K_1, \forall r \in S(x_1)x_2(x_1) \leq 0. \\
\tag{**}
\]
Proof. If (**) applies, then it follows from the proof of 3.3.1 that an $I_2$ exists so that (*) holds.

On the other hand, suppose that for $I_2 \in \mathcal{R}_n(A)$ (*) applies.

If $\nu_t = 0$ for all $t \in I_1 | K_1$, then (**) follows trivially.

Assume that there is a $t \in I_1 | K_1$ for which $\nu_t \neq 0$.

Let $I_1 | I_2 = \{s\}$, then $s \in K_1$, since $\rho(A(I_2), b(I_2)) > \rho(A(I_1), b(I_1))$.

Because $\nu_t \neq 0$ and $\nu_s \neq 0$, $t$ as well as $s$ are contained in $K_2$.

Now we have

$$x_1 \in S(K_2) \Rightarrow r_t(x_1) \nu_t = r_s(x_1) \nu_s = -r_s(x_1)$$

or

$$r_t(x_1) \nu_t r_s(x_1) < 0,$$

and (**) is obtained. \qed

Remark.

If for any $i \in I_1 | K_1$, $\nu_i r_t(x_1) r_s(x_1) \leq 0$, then there is an $I_2$ so that

$$\begin{cases} x_1 \in S(K_2) \\ \rho(A(I_2), b(I_2)) > \rho(A(I_1), b(I_1)) \end{cases},$$

as well as an $I_2$ so that

$$\begin{cases} x_1 \in S(K_2) \\ \rho(A(I_2), b(I_2)) = \rho(A(I_1), b(I_1)) \end{cases},$$

provided that for some $t \in I_1 | K_1$, $\nu_t r_t(x_1) r_s(x_1) < 0$. \qed

Remark.

If $\rho(A(I_1), b(I_1)) = 0$, then we choose $s$ so that

$$r_t(x_1) \frac{\nu_s}{\lambda_s} = \max_{i \in K_1} \left( r_s(x_1) \frac{\nu_i}{\lambda_i} \right).$$

With this choice of $s$ it readily follows that $\rho(A(I_2), b(I_2)) \neq 0$.

Remark.

If $s$ is so chosen that $x_1 \notin S(K_2)$, then

$$\rho(A(I_2), b(I_2)) \geq \rho(A(I_1), b(I_1))$$
as well as
\[ \rho(A(I_2), b(I_2)) < \rho(A(I_1), b(I_1)) \]
is possible.

Should one wish to apply the classical algorithm to a system not satisfying the Haar condition, then the criterion that determines what equation will be exchanged should be adapted. It is obvious that we still want that \( x_1 \in S(K_2) \), since otherwise the deviation may decrease.

On the other hand, we know that with this criterion the deviation may increase as well as remain stationary, even though an \( I_2 \) exists for which the deviation does increase.

Suppose the criterion is adapted as follows:

"Choose \( I_2 \) so that

(i) \( x_1 \in S(K_2) \)

(ii) \( \rho(A(I_2), b(I_2)) \) is, if possible, greater".

We saw that this criterion indeed determines an \( I_2 \), although it does not do so uniquely. Furthermore, we saw that it is not always possible to determine \( I_2 \) satisfying (i) for which the deviation does increase. This implies that the possibility of "cycling" is not excluded.

An example of cycling:

Consider the system of linear equations

\[
\begin{bmatrix}
1 & 0 \\
0 & 1 \\
1 & 0 \\
0 & 1
\end{bmatrix} x = \begin{bmatrix}
0 \\
2 \\
4 \\
3
\end{bmatrix}.
\]

By simple direct analyses it follows that \( \{1,3\} \) is the index-set corresponding to the maximal cadre matrix, that the deviation is 2 and that

\[ T = \{ x \mid (^2_1) \leq x \leq (^2_4) \}. \]

Choose as starting reference \( I_1 = \{1,2,3\} \).

\[ \rho(A(I_1), b(I_1)) = 2 \] and an extremal minimax solution is: \( x_1 = (^2_0) \).

\[ \|x(1)\|_\infty = 3, \] and since we know that the deviation is already at its maximal value, \( A(I_2) \) certainly contains the maximal cadre matrix so \( I_2 \) must be: \( \{1,3,4\} \). An extremal minimax solution is: \( x_2 = (^2_5) \).

\[ \|x(2)\|_\infty = 3 \] and by the same argument as above, \( I_3 \) must be: \( \{1,2,3\} = I_1 \).
3.4. In this part an algorithm, due to Descloux, will be discussed, which is a variant of the classical algorithm and which may be employed even though the Haar condition is not satisfied.

If a reference $A(I)$ does not satisfy the Haar condition, this is equivalent to the event that, if $\lambda^T A(I) = \theta^T$ with $\lambda \neq \theta$, some of the $\lambda_i$ are equal to zero. Obviously, an arbitrary perturbation of the elements of $A(I)$ will, generally, cause the perturbed $\lambda_i$ all to be distinct from zero and, consequently, the perturbed $A(I)$ will satisfy the Haar condition. This might be compared with the idea of Charnes in linear programming, to perturb the right-hand vector in case of degenerate bases.

So, when dealing with minimax problems where the matrix of coefficients does not satisfy the Haar condition, it seems to be a good idea to perturb the matrix in an artificial way so that the classical algorithm may be applied legally. A clear disadvantage of this method would be that, in advance, it is hard to tell how much the ultimately found "solution" will deviate from the set $T(A, b)$, and, moreover, the perturbation really has to be introduced. However, necessary for legal application of the classical algorithm is not that the matrix of coefficients satisfies the Haar condition, but that any reference occurring in some iteration satisfies the condition. So a perturbation, if necessary, of the current reference, would make the application legal. Also here the question can be asked what the influence of accumulative perturbations will be on the "solution". However, we shall see that it is possible to simulate a sequence of iterations that may be generated by applying the classical algorithm to a perturbed minimax problem, without really perturbing the matrix of coefficients, which also implies that it is possible to overcome the effect of accumulating perturbations.

First of all let us prove that a reference can be perturbed in such a way that it satisfies the Haar condition and, that, moreover, any other reference having the same cadre matrix also satisfies the condition. We need the following lemma.

3.4.1. Lemma.

Let $C_1, \ldots, C_{p+1}$ be subspaces of $\mathbb{R}^n$ such that $C_1 + \ldots + C_{p+1}$. Then there is an $i \in \{1, \ldots, p\}$ so that $C_{p+1} \subseteq C_i$. 

Let $C_1, \ldots, C_{p+1}$ be subspaces of $\mathbb{R}^n$ such that $C_1 + \ldots + C_{p+1}$. Then there is an $i \in \{1, \ldots, p\}$ so that $C_{p+1} \subseteq C_i$. 

Let $C_1, \ldots, C_{p+1}$ be subspaces of $\mathbb{R}^n$ such that $C_1 + \ldots + C_{p+1}$. Then there is an $i \in \{1, \ldots, p\}$ so that $C_{p+1} \subseteq C_i$. 

Let $C_1, \ldots, C_{p+1}$ be subspaces of $\mathbb{R}^n$ such that $C_1 + \ldots + C_{p+1}$. Then there is an $i \in \{1, \ldots, p\}$ so that $C_{p+1} \subseteq C_i$.
Proof. Let $p > 1$.

Remove those $C_i$ for which holds that

$$C_i \cap C_{p+1} \subseteq \bigcup_{j=1}^{p} (C_j \cap C_{p+1})_{j \neq i}.$$

Suppose that $C_1, \ldots, C_q$ remain, then

$$C_{p+1} \subseteq \bigcup_{j=1}^{q} C_i.$$

We shall show that the assumption that $q > 1$ leads to a contradiction.

We have, if $q > 1$,

$$C_{p+1} \cap C_1 \neq \bigcup_{j=1}^{q} (C_j \cap C_{p+1}), \quad C_{p+1} \cap C_2 \neq \bigcup_{j=1}^{q} (C_j \cap C_{p+1}),$$

and, consequently, $x_1$ and $x_2$ exist so that

$$\begin{cases} x_1 \in C_{p+1} \cap C_1 \text{ and } x_1 \notin C_j, \\ x_2 \in C_{p+1} \cap C_2 \text{ and } x_2 \notin C_j \end{cases} \quad (\ast)$$

Let $V$ be the one-dimensional hyperplane (line) containing $x_1$ and $x_2$, $C_{p+1}$, being a subspace, contains $V$ and so

$$V \subseteq \bigcup_{i=1}^{q} C_i.$$

It is obvious that there must be a $C_i$ containing $V$, since there must be a $C_i$ containing two points of $V$.

This is contrary to $(\ast)$ and so $q$ must be equal to one, which concludes the proof.

3.4.2. Corollary.

Let $C_1, \ldots, C_p$ be proper subspaces of $\mathbb{R}^n$.

Then $\bigcup_{i=1}^{p} C_i$ is a proper, non-trivial subset of $\mathbb{R}^n$. 

\[\square\]
Proof. Let \( C_{p+1} := \mathbb{R}^n \) and apply 3.4.1. 

3.4.3. Theorem.

Let \( A \in M_{m \times n} \) with \( m > n \).

Let \( K \in \mathcal{H}_k(A) \) with \( k < n \) and let \( p \in K \).

Let, if \( u_p \neq 0 \) is some vector in \( \mathbb{R}^n \), \( A^c \) be the matrix obtained from \( A \) by replacing \( a_p \) by

\[
\frac{a^c}{p} := \frac{a}{p} - \varepsilon u_p,
\]

where \( \varepsilon \) is a scalar.

Then there is a vector \( u_p \in \mathbb{R}^n \) and there is an \( \varepsilon_0 > 0 \) so that for any \( I \in \mathcal{R}_n(A) \) with \( I \supseteq K \) applies:

\( A^c(I) \) satisfies the Haar condition, provided that \( 0 < |\varepsilon| < \varepsilon_0 \).

Proof. Let \( A \neq 0 \) so that \( \sum_{i \in K} \lambda_i a_i = 0 \) and define for \( i \in L|K|, \lambda_i := 0 \).

Let \( \mathcal{P}(K) := \{ I \in \mathcal{R}_n(A) \mid I \supseteq K \} \).

\( \mathcal{P}(K) \) is finite.

Let

\[
W := \bigcap_{I \in \mathcal{P}(K)} \{ \mathbb{R}^n \mid \{ \text{span of (n-1) linearly independent rows of } A(I) \} \}
\]

From corollary 3.4.2 it follows that \( W \) is open and non-trivial.

Let \( u \neq 0 \in W \).

For any \( I \in \mathcal{P}(K) \) uniquely determined scalars \( \delta_j(I) \) exist so that

\[
u = \sum_{j \in I\{p\}} \delta_j(I) a_j.
\]

Since \( u \in W \), we have

\[
\forall I \in \mathcal{P}(K), \text{ for } j \in I\{p\}, \delta_j(I) \neq 0.
\]

Let \( \varepsilon \) be a scalar and \( A^c \) be obtained from \( A \) by replacing \( a_p \) by \( a^c_p := a_p - \varepsilon u \).

We obtain

\[
\forall I \in \mathcal{P}(K), A^c(I) \text{ is a reference},
\]
and, moreover,
\[ \forall I \in \mathcal{P}(K), \lambda_{p}^\epsilon = \lambda_{p} (a_p - \epsilon u) \]
\[ = \sum_{i \in I \setminus K} -\epsilon \lambda_{p} \delta_i(I) a_i + \sum_{i \in K \setminus \{p\}} -(\lambda_i - \epsilon \lambda_{p} \delta_i(I)) a_i \]
\[ = -\sum_{i \in I \setminus \{p\}} \lambda_i^\epsilon a_i, \]

with
\[ \lambda_{p}^\epsilon = \lambda_{p} \]
\[ \lambda_i^\epsilon = \lambda_i + \epsilon \delta_i(I), \text{ for } i \in K \setminus \{p\}, \]
\[ \lambda_i^\epsilon = \epsilon \lambda_{p} \delta_i(I), \text{ for } i \in I \setminus K. \]

For any \( I \in \mathcal{P}(K) \) an \( e_0(I) \) exists so that for \( 0 < |\epsilon| < e_0(I), \lambda_i^\epsilon \neq 0. \) Consequently, we have
\[ \forall I \in \mathcal{P}(K), I \in \mathcal{A}_n(A^\epsilon), \text{ provided that } 0 < |\epsilon| < e_0, \]
with \( e_0 = \min_{I \in \mathcal{P}(K)} e_0(I), \) which completes the proof.

3.4.4. Definition.

If \( u_p \) satisfies the conditions of theorem 3.4.3, then \( u_p \) will be called a genuine perturbation of \( a_p \).

3.4.5. The algorithm of Descloux.

Let \( A \in \mathcal{M}_{m,n}^n \) with \( m > n \) and \( b \in \mathbb{R}^n \), then

(0) \( j := 0, \)

(1) choose \( I \in \mathcal{P}_n(A), \)

(2) determine \( \lambda \neq 0 \) so that
\[ \sum_{i \in I} \lambda_i a_i = \theta, \]
and compute
\[ \rho(I) := (-\sum_{i \in I} \lambda_i a_i)(\sum_{i \in I} |\lambda_i|)^{-1}, \]

after which, if necessary, \( \lambda \) is adapted so that \( \rho(I) \geq 0, \)
(3) determine the cadre \( K \) of \( I \),
\[ K := \{ i \mid i \in I, \lambda_i \neq 0 \}, \]
and choose coefficients \( \gamma_i \neq 0 \) for \( i \in I \setminus K \).

(4) determine
\[ \text{for } i \in K, \quad r_i := \rho(I) \text{ sign } (\lambda_i), \]
\[ \text{for } i \in I \setminus K, \quad r_i := \rho(I) \text{ sign } (\gamma_i), \]

(5) compute the solution \( z \) of
\[ A(I)x = b(I) + \zeta(I), \]

(6) determine \( \ell \in L|I \) so that \( |r_\ell(x)| \) is maximal, if \(|r_\ell(x)| \leq \rho(I)\)
then \( x \) is a minimax solution of \( A(z) = b \), otherwise,

(7) compute \( \nu \) so that
\[ a_{\ell} = \sum_{i \in I} \nu_i a_i, \]

(8) determine \( s \in K \) and \( M \) so that
\[ M := r_\ell(x) \frac{\nu_i}{\lambda_i} = \max \left( r_\ell(x) \frac{\nu_i}{\lambda_i} \right), \]
determine \( t \in I \setminus K \) and \( N \) so that
\[ N := r_\ell(x) \frac{\nu_i}{\gamma_i} = \max \left( r_\ell(x) \frac{\nu_i}{\gamma_i} \right), \]
if \( K = I \), then \( N := 0 \),

(9) if \( N \leq 0 \), then \( I := (I \setminus \{ s \}) \cup \{ \ell \}, \quad j := j + 1 \) and return to (2),

(10) if \( N < 0 \), then \( I := (I \setminus \{ t \}) \cup \{ \ell \}, \quad j := j + 1 \),
for \( i \in (I \setminus K) \setminus \{ t \} \), if \( \gamma_i - \frac{\nu_t}{\gamma_t} \nu_i \neq 0 \), then \( \gamma_i := \gamma_i - \frac{\nu_t}{\gamma_t} \nu_i, \)
\[ \gamma_t := \frac{\nu_t}{\gamma_t}, \]

(11) return to (4).
Remark.

As is the case for the classical algorithm, also here in each iteration a subsystem \( A(I_j)x = b(I_j) \) is considered.

In 3.4.6 it will be shown that for any \( j \) \( A(I_j) \) is a reference. For the corresponding minimax problem the deviation \( \rho(I_j) \) and a minimax solution \( x_j \) are computed. The difference between two consecutive iterations is that \( A(I_j) \) and \( A(I_{j+1}) \) are equal except for one row. The result of the adjustment of the \( \gamma_i \) in (10) is that for consecutive minimax solutions \( x_j \) and \( x_{j+1} \) holds that

\[
\text{for } i \in I_{j+1}, \text{ sign } r_i(x_j) = \text{sign } r_i(x_{j+1}),
\]

which is compatible with the classical algorithm.

If the algorithm of 3.4.5 is applied to a system satisfying the Haar condition, then in each iteration we have that \( N = 0 \) and so the same exchange-criterion is used as in the classical algorithm.

That indeed a sequence of iterations is simulated that may be generated by applying the classical algorithm to a perturbed system, will become clear in the next pages.

The finiteness of the classical algorithm followed from the fact that in each iteration the deviation did increase. However, the proof of the finiteness of the algorithm of Descloux is more complicated, since only can be stated that the deviation does not decrease.

3.4.6. Theorem.

The algorithm of Descloux, applied to \( Ax = b \) with \( A \in \mathbb{R}^{n \times m} \) and \( m > n \), supplies a sequence of iterations for which holds that the corresponding sequence of deviations is non-decreasing.

If during a number of iterations the deviation is stationary, then the corresponding references all have the same cadre matrix.

Proof. In any iteration there are two possibilities: either \( N \leq 0 \) and so

\[
\forall i \in I, \nu_i r_i(x) r_i(x) \leq 0,
\]

or \( N > 0 \) and so

\[
\exists p \in I, \nu_p r_p(x) r_p(x) > 0.
\]
According to corollary 3.3.1 in the first case a reference $I'$ with cadre $K'$ exists so that

$$\begin{align*}
\mu(I'|I) = \mu(I|I') = 1 \text{ and } x \in S(K'), \\
\rho(A(I'),b(I')) > \rho(A(I),b(I)).
\end{align*}$$

(1) (2)

It is easy to verify that in the next iteration the algorithm supplies a reference $I'$ satisfying (1) and (2). In the second case no $I'$ exists so that (1) and (2) are true, but, according to theorem 3.3.1, an $I'$ exists so that (1) is satisfied and

$$\rho(A(I'),b(I')) = \rho(A(I),b(I)),$$

(3)

and, moreover, $I$ and $I'$ have the same cadre.

It is easy to verify that the algorithm indeed supplies a reference $I'$ satisfying (1), (3) and (4).

The proof is concluded by the remark that $A(I_1)$ is a reference so that as a consequence of the foregoing all $A(I_j)$ are references.

3.4.7. Definition.

A sequence of iterations supplied by the algorithm of Descloux is called an étappe, if it consists of more than one iteration and the corresponding sequence of deviations is stationary.

If, when employing the algorithm of Descloux, in some iteration the deviation does increase, then, as a consequence of 3.4.6, none of the previous references can become current again. So, in order to prove the finiteness of the algorithm, it is sufficient to prove that any étappe is finite.

In the first reference of an étappe for $i \in I|K$ scalars $\gamma_i \neq 0$ are chosen.

We shall prove that an étappe is finite, provided that the scalars $\gamma_i$ satisfy the following condition:

$$\begin{align*}
\text{Let } A(I) \text{ be the first reference of an étappe, and let } A(K) \text{ be its cadre matrix.} \\
\text{The scalars } \gamma_i \neq 0 \text{ are, for } i \in I|K, \text{ so chosen that for any } p \in K, \text{ an } u^*_p \in \text{Span } \{a_i | i \in K\} \text{ exists so that} \\
u_p := u^*_p + \sum_{i \in I|K} \gamma_i a_i \\
is a genuine perturbation of } a_p.
\end{align*}$$

(D)
Remark.

According to theorem 3.4.4 it is indeed possible that the initial scalars $y_i$ satisfy condition (D). Moreover, if for some $p \in K$ a $u_p$ exists, then for any $p \in K$ a $u_p$ exists.

3.4.8. Theorem.

Let the algorithm of Descloux be applied to a system of linear equations $Ax = b$ with $A \in \mathbb{M}_{m,n}$, $m > n$.

Suppose that an étappe occurs fixed by

$$ (I_1, x_1), (I_2, x_2), \ldots, $$

where the references $A(I_j)$ have the common cadre matrix $A(K)$ and the common deviation $p(K)$.

Let the scalars $y_i^{(1)}$ that are chosen in the first iteration of (1) satisfy the condition (D) so that for some $p \in K$, $u_p$ is the corresponding genuine perturbation of $a_p$. Then holds that the adjusted scalars $y_i^{(j)}$ also satisfy condition (D) with the same genuine perturbation $u_p$.

Proof. Let $\lambda \neq 0$ so that $\lambda^T A(K) = 0^T$. Let $\varepsilon$ be a small scalar that is unequal to zero and let $A^\varepsilon$ be obtained from $A$ by replacing $a_p$ by $a_p - \varepsilon u_p$.

We have, for any $I_j \in (A)$, provided that $|\varepsilon|$ is small enough,

$$ \lambda^T A^\varepsilon(I_j) + \varepsilon \lambda u_p^T = 0^T, $$

and, by 3.4.4,

$$ I_j \in \mathcal{H}_n(A^\varepsilon). $$(3)

Since the scalars $y_i^{(1)}$ satisfy condition (D) with corresponding genuine perturbation $u_p$, we obtain

$$ u_p - \sum_{i \in I_1}^{L} y_i^{(1)} a_i \in \text{Span} \{a_i | i \in K\}. $$

We have to prove that, for any $I_j \in (1)$,

$$ u_p - \sum_{i \in I_j}^{L} y_i^{(j)} a_i \in \text{Span} \{a_i | i \in K\}. $$

(4)
By simple calculations, using 3.4.5, we obtain

\[ u_p = \sum_{i \in I_1} \gamma_i^{(1)} a_i + u^* \]

\[ = \sum_{i \in I_1} \left( \gamma_i^{(1)} - \frac{\nu_i}{\nu_t} \gamma_t^{(1)} \right) a_i + \frac{\gamma_t^{(1)}}{\nu_t} a_k + v^*, \]

where \( u^* \) and \( v^* \) both belong to Span \( \{ a_i | i \in K \} \).

Substituting this result in (2), it follows from (3) and 1.1.4 that

for \( i \in I_1 | K | \{ t \} \), \( (\gamma_i^{(1)} - \frac{\nu_i}{\nu_t} \gamma_t^{(1)}) \neq 0 \),

so that by 3.4.5

\[ u_p = \sum_{i \in I_2} \gamma_i^{(2)} a_i \in \text{Span} \{ a_i | i \in K \}, \]

and the proof is completed by an inductive argument. \( \square \)

3.4.9. Corollary.

Let \( I_j \) be the reference occurring in some iteration of (1). Then

for \( i \in I_j | K | \), \( \gamma_i^{(j)} \neq 0 \). \( \tag{5} \)

3.4.10. Corollary.

The étappe fixed by (1) is finite.

Proof. Consider the sequence

\[ (I_1, x_1^c), (I_2, x_2^c), \ldots \] \( \tag{6} \)

where \( x_j^c \) is the minimax solution of \( A^c(I_j)x = b(I_j) \).

Let the corresponding sequence of deviations be

\[ \rho_1^c, \rho_2^c, \ldots \] \( \tag{7} \)
As a consequence of (2) and (4) we obtain for any $I_j \in (1)$,

$$
\sum_{i \in K} (\lambda_i + \epsilon(\epsilon))a_{i1}^\epsilon + \sum_{i \in I_j} \epsilon \gamma_j a_{i1}^\epsilon = \sum_{i \in I_j} \lambda_i a_{i1}^\epsilon = 0. \tag{8}
$$

Using 2.2.1 we obtain, as a consequence of (8),

$$
\lim_{\epsilon \to 0} \rho_j^\epsilon = \rho(K). \tag{9}
$$

As a consequence of (8), (9), 3.4.5 and 2.2.1 we have, provided that

$$
\text{sign } \epsilon = \text{sign } \lambda_p,
$$

for $i \in I_j$, $\text{sign } r_i(x_j) = \text{sign } r_i(x_j^\epsilon), \tag{10}
$$

or, equivalently,

$$
\text{for } i \in I_j \setminus \{p\}, \; a_i^T(x_j^\epsilon - x_j) = (\rho_j^\epsilon - \rho(K)) \text{ sign } r_i(x_j),
$$

which, using (9), implies that

$$
\lim_{\epsilon \to 0} x_j^\epsilon = x_j, \tag{11}
$$

provided that $\text{sign } \epsilon = \text{sign } \lambda_p$.

Let us suppose that $I_{j+1} \in (1)$. This implies that

$$
|r_\lambda(x_j)| > \rho(K) \text{ and } N > 0, \tag{12}
$$

where $\lambda$ and $N$ are defined as in 3.4.5.

As a result of (11) and (12) we obtain, provided that

$$
|\epsilon| \text{ is small enough and } \text{sign } \epsilon = \text{sign } \lambda_p, \tag{13}
$$

$$
|r_\lambda(x_j^\epsilon)| > \rho_j^\epsilon. \tag{14}
$$

Applying theorem 3.2.1 and the first remark following 3.2.1 to the system $A^\epsilon(I_j \cup \{\ell\})x = b(I_j \cup \{\ell\})$, we obtain that a reference $I_j' := (I_j \setminus \{s\}) \cup \{\ell\}$ exists, where $s$ is determined by

$$
r_\lambda(x_j) \frac{v_s}{\epsilon \lambda \gamma_j(\epsilon)} = \max_{i \in I_j} \left( r_\lambda(x_j^\epsilon) \frac{v_i}{\epsilon \lambda \gamma_j(\epsilon)} \right),
$$

so that $\rho(A^\epsilon(I_j'), b(I_j')) > \rho_j^\epsilon$. 


If (13) is true, we obtain, since $N \neq 0$,

$$r_j(x_j^\varepsilon) \frac{\nu_s}{\varepsilon \lambda_j^{\varepsilon} \gamma_s^{(j)}} = \max_{i \in I|K} \left( r_i(x_i^\varepsilon) \frac{\nu_i}{\lambda_i^{\varepsilon}} \right),$$

and, consequently, it follows from 3.4.5 that $s = t$, or, equivalently,

$$I_j^\varepsilon = I_{j+1}^\varepsilon,$$

which implies

$$p_{j+1}^{\varepsilon} > p_j^{\varepsilon}. \tag{15}$$

The consequence of (15) is that the sequence (6) must be finite, and, therefore, the sequence (1) is finite.

3.4.11. Theorem.

The algorithm of Descloux applied to a system $Ax = b$, with $A \in \mathbb{R}^{m \times n}$ and $m > n$, supplies after finitely many iterations a minimax solution of $Ax = b$, provided that in any étappe the scalars $\gamma_i$ satisfy the condition (D).

Proof. It follows from 3.4.8 and 3.4.10 that the sequence of iterations, supplied by the algorithm, is finite. The consideration that the last computed minimax solution is a minimax solution of the entire system $Ax = b$, concludes the proof.

Remark.

In 3.4.11 the reserve is made that the choice of the scalars $\gamma_i$ in the first iteration of an étappe satisfies condition (D).

If so, then in none of the iterations, as a consequence of 3.4.9, the adjusted scalars $\gamma^{(j)}_i$ are equal to zero, or, equivalently, in any iteration the value of $N$ is attained for a unique $t \in I|K$. It is quite clear that it is very unlikely, that an arbitrary choice of the scalars will not satisfy condition (D), but it is possible that the condition is not satisfied. The obvious remedy in this case is to adapt the $\gamma_i$ that would become equal to zero so that the value of $N$ is attained for only one $t \in I|K$. Indeed, this is what is done in the algorithm. One may consider the adapted $\gamma_i$ as corresponding with the start of a new étappe. That the phenomenon occurs again is even more unlikely and we may state:

"The chance that the algorithm of Descloux is infinite is equal to zero".
Remark.

One might expect that to any extremal minimax solution a sign-distribution of the scalars $y_i$, chosen in the last étappe, corresponds. If the rank of $A(K)$ equals $K$, then $2^{n-k}$ sign-distributions are possible. However, the number of extremal vectors of $T(A,b)$ may exceed $2^{n-k}$ and this implies that the sizes of the $y_i$ will also determine what extremal vector will finally be found.

3.5. One may wonder if the manipulation of the scalars $y_i$ in the algorithm of Descloux is indeed necessary.

Let us consider the exchange in an étappe more closely. We have, if $I_1$ and $I_2$ determine any two consecutive references in an étappe,

$$I_2 = (I_1 \cup \{t\}) \cup \{i\},$$

where $i \in L|I_1$ so that

$$|r_{\iota}(x_1)| > \rho(K),$$

and where $t \in I_1|K$ so that

$$r_{\iota}(x_1) v_{\iota} = \max_{i \in I_1|K} \left( r_{\iota}(x_1) \frac{v_i}{y_i} \right).$$

It is easy to verify that also holds

$$v_{\iota} r_{\iota}(x_1) r_{\iota}(x_1) > 0. \quad (1)$$

Let us investigate what happens if we would take

$$I_2' = (I_1 \cup \{p\}) \cup \{i\},$$

where $p \in I_1|K$ so that

$$v_{\iota} r_{\iota}(x_1) r_{\iota}(x_1) < 0, \quad (2)$$

premising that indeed such a $p$ exists.

In this case, too, we have

$$\rho(I_2') = \rho(K),$$

and we can arrange that

for $i \in I_2'$, sign $r_i(x_1) = \text{sign} r_i(x_2)$,
where $\tilde{x}_2$ denotes an extremal minimax solution of $A(I_2')\tilde{x} = b(I_2')$.

Considering the residual in the equation leaving $I_1$ we obtain

$$r_p(x_2) = |a^T_p x_2 - b_p| = |(a^T_p x_2 - a^T_p x_1) + (a^T_p x_1 - b_p)|$$

$$= \left|\frac{1}{v_p} (r_p(x_2) - r_p(x_1)) + \rho(K) \; \text{sign} \; r_p(x_1)\right|$$

$$= \frac{\text{sign} \; r_p(x_1) \; \; \text{sign} \; r_p(x_1)}{v_p} (\rho(K) - |r_p(x_1)|) + \rho(K),$$

and, by (2),

$$|r_p(x_2)| > \rho(K).$$

Consequently, with this choice of the equation to be exchanged, it is possible that the algorithm cycles, namely, if holds

$$|r_p(x_2)| = \max_{i \in L | I_2'} |r_i(x_2)|.$$ 

On the other hand, if the equation to leave the reference satisfies (1), similar calculations lead to

$$r_t(x_2) = \left[\frac{\text{sign} \; r_t(x_1) \; \; \text{sign} \; r_t(x_1)}{v_t} (\rho(K) - |r_t(x_1)|) + \rho(K)\right] \text{sign} \; r_t(x_1),$$

and thus,

$$r_t(x_2) < \rho(K) \; \text{sign} \; r_t(x_1),$$

and one may verify that this, for the simple case that $\mu(L) = n+2$, prohibits cycling. Considering this and the fact that, in an étappe, the algorithm of Descloux exchanges an equation satisfying (1), we venture the following conjecture:

If the classical algorithm, in order to make it general applicable, is so adapted that in case there is no $I_2$ for which both $\rho(I_2) > \rho(I_1)$ and $\tilde{x}_1 \in S(K_2)$, the equation to be exchanged satisfies (1), then this variant will supply an extremal minimax solution of $A\tilde{x} = b$ in a finite number of steps.

Remark.

It is easy to verify that this variant, too, supplies a non-decreasing sequence of deviations.
3.5.1.
An elaboration of this algorithm might be the following

Let $A \in \mathcal{M}^{n \times n}$, $m > n$ and $b \in \mathbb{R}^m$, then

1. $j := 0,$
2. choose $I \in \mathcal{R}_n(A),$
3. determine $\lambda \neq \emptyset$ so that
   $$\sum_{i \in I} \lambda_i a_i = \emptyset,$$
   and compute
   $$\rho(I) := (-\sum_{i \in I} \lambda_i b_i)(\sum_{i \in I} |\lambda_i|)^{-1},$$
   after which, if necessary, $\lambda$ is adapted so that $\rho(I) \geq 0,$
4. determine the cadre $K$ of $I$,
   $$K := \{i | i \in I, \lambda_i \neq 0\},$$
   and choose scalars $\sigma_i \neq 0$ for $i \in I|K$ so that $|\sigma_i| = 1,$
5. determine
   for $i \in K$, $r_i := \rho(I) \text{ sign } (\lambda_i),$
   for $i \in I|K$, $r_i := \rho(I) \sigma_i,$
6. compute the solution $x$ of
   $$A(I)x = b(I) + \varepsilon(I),$$
7. determine $\ell \in L|I$ so that $|r_{\ell}(x)|$ is maximal, if $|r_{\ell}(x)| \leq \rho(I)$
   then $x$ is the desired minimax solution of $Ax = b$, otherwise,
8. compute $\nu$ so that
   $$\sigma_{\ell} = \sum_{i \in I} \nu_i a_i,$$
   if for some $t \in I|K$ it appears that
   $$\nu_t r_{\ell}(x) \sigma_i > 0,$$
   then $I := (I\{|t\}) \cup \{\ell\}, j := j+1, \sigma_t := \text{ sign } r_{\ell}(x)$ and return to
   (4), otherwise,
(9) determine \( s \in K \) and \( M \) so that

\[
M := r^*_k(x) \frac{\nu_s}{\lambda_s} = \max_{i \in K} (r^*_k(x) \frac{\nu_i}{\lambda_i}),
\]

\( I := (I|\{s\}) \cup \{\ell\}, \ j := j+1 \) and return to (2).

\[ \square \]

Remark.

The manipulation of the scalars \( \gamma_i \) has totally disappeared!

We have not succeeded in proving the finiteness of the algorithm for arbitrary \( A \in \mathcal{M}_{m,n} \) with \( m > n \).

We were able to prove that an étappe, for which \((n-k)\) equals 1 or 2, must end and we were also able to prove the finiteness for \( A \in \mathcal{M}_{n+2,n} \).

We found that it is essential that \(|r^*_k(x)|\) should be the maximal residual.

Omitting this condition may cause the cycling of the algorithm as is illustrated in figure 2 for an étappe with \((n-k) = 2\). The plane of drawing corresponds with the set of minimax solutions of the cadre system, whereas the other equations cut slabs out of this plane. Suppose that the algorithm is applied with starting point 1.
This cycle is possible if and only if

(i) VI has maximal residual with respect to 1
(ii) V " " " " " " 2
(iii) IV " " " " " " 3
(iv) III " " " " " " 4
(v) II " " " " " " 5
(vi) I " " " " " " 6.
Let $d(\text{II}), d(\text{IV})$ and $d(\text{VI})$ denote the width of the slabs II, IV and VI. From the figure it follows that

by (i), $d(\text{VI}) < d(\text{IV})$,
by (iii), $d(\text{IV}) < d(\text{II})$,
and by (v), $d(\text{II}) < d(\text{VI})$,

but this implies $d(\text{VI}) < d(\text{VI})$. So, if at any iteration the equation to enter the reference is the one of maximal residual, then this particular cycle is not possible.

3.5.2. In a paper of Osborne and Watson, 1968, the equivalence of the classical algorithm with the simplex algorithm, applied to the dual of the linear programming formulation of the minimax problem, is demonstrated. The dual of the L.P.-formulation for $\min \|Ax - b\|_\infty$ is:

\[
\begin{bmatrix}
A^T & -A^T \\
e^T & e^T
\end{bmatrix}
\begin{bmatrix}
y \\
1
\end{bmatrix} = \begin{bmatrix}0 \\
1
\end{bmatrix},
\]

$\mathbf{w} \geq 0$,

\[
\max (b^T, -b^T)\mathbf{w},
\]

where $\mathbf{e}$ denotes the vector $(1, 1, \ldots, 1)$.

Considering the simplex algorithm applied to this L.P.-problem, Osborne and Watson remark that an optimal basic solution exists and that an optimal basis does not contain duplicated columns of $A^T$ and, consequently, that after a number of iterations no basis contains duplicated columns of $A^T$. Furthermore, they show that the equation to enter the basis is the same for the classical algorithm and the simplex algorithm and that, provided that $A$ satisfies the Haar condition the equation to leave the basis is also in both cases the same. However, the simplex algorithm can be applied even if $A$ does not satisfy the Haar condition, although then the possibility of cycling is not excluded. We shall show that in that case the simplex algorithm is precisely identical to the algorithm proposed in 3.5.1. The only thing that is left to prove is that in both cases the equation to be exchanged is the same.

The simplex algorithm exchanges the equation for which applies

\[
p := \frac{w_i}{y_i} = \min \frac{w_i}{y_i} \mid i \in I, y_i > 0\},
\]
where $y_i$ equals $\nu_i \text{sign } r_i(x) \text{sign } r_{\lambda}(x)$ and $w_i$ equals $|\lambda_i|$, so $p \geq 0$.

If $p = 0$, the equation to be exchanged is one for which $|\lambda_i| = 0$ and $\nu_i \text{sign } r_i(x) \text{sign } r_{\lambda}(x) > 0$.

If $p > 0$, the equation to be exchanged is one for which $|\lambda_i| \neq 0$ and $\text{sign } r_{\lambda}(x) \cdot \frac{\nu_i}{\lambda_i}$ is maximal.

And these two conditions are precisely identical to the criterion used by 3.5.1.
4. In this section we shall consider the problem to find an $x \in \mathbb{R}^n$ so that

$\begin{align*}
\begin{cases}
A_1 x = b_1 , \\
\|A_2 x - b_2\|_\infty \text{ is minimal}
\end{cases}
\end{align*}$

where $A_1 \in \mathcal{M}_{r,n}^{r}$ and $(A_2) \in \mathcal{M}_{m,n}^{n}$, $m > n > r$.

This is a generalized form of the ordinary minimax problem with some more practical importance.

Let us consider, for instance, a problem due to Stiefel [13]:

Given $f(x)$ and $f'(x)$ at the points $x_1, \ldots, x_n$.

Find a polynomial $\phi(x) = \sum_{j=0}^{n+r} a_j x^j$ of maximal degree $(n+r)$, $0 \leq r < n$, so that

for $i = 1, \ldots, n$, $\phi(x_i) = f(x_i)$,

$max |\phi'(x_i) - f'(x_i)|$ is minimal.

Henceforth, the generalized problem will be denoted as GMP and the ordinary minimax problem will be denoted as OMP.

4.1. The GMP can be reduced to an OMP.

Let us consider the following GMP:

Find $x_1, x_2 \in \mathbb{R}^n$ so that

$\begin{align*}
\begin{cases}
(A_{11}|A_{12}) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = b_1 , \\
\| (A_{21}|A_{22}) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - b_2 \|_\infty \text{ is minimal,}
\end{cases}
\end{align*}$

(1)

with $A_{11} \in \mathcal{M}_{r,r}^{r}$ and $(A_{11}|A_{12}) \in \mathcal{M}_{m,n}^{n}$, $m > n > r$.

This problem is equivalent with the OMP:

Find $y$ so that

$\| (A_{22} - (A_{22}^{-1}A_{11}^{-1}A_{12})) y - (A_{22}^{-1}A_{21}^{-1}b_1) \|_\infty$ is minimal,

(2)

with $A_{11} \in \mathcal{M}_{r,r}^{r}$ and $(A_{11}|A_{12}) \in \mathcal{M}_{m,n}^{n}$, $m > n > r$. 

Indeed, if $\mathbf{x}_2$ is a solution of (1), then $\mathbf{y} := \mathbf{x}_2$ is a solution of (2) and, conversely, if $\mathbf{y}$ is a solution of (2) then

$$
\mathbf{x}_2 := (A_1^{-1}b_1 - A_1^{-1}A_2\mathbf{y}) / \mathbf{y}
$$

is a solution of (1).

This implies that it is possible to translate the results of the sections 2 and 3 for the generalized problem, and that the finding of a solution of (1) may be achieved by applying the algorithm of Descloux to the OMP (2). However, a direct treatment of the GMP is more attractive. We shall see that former theorems are valid with minor alterations and, moreover, that only a slight change in the algorithms given for the OMP will make them applicable to the GMP.

4.2. In the next pages we shall study the GMP:

$$
\begin{align*}
\begin{cases}
A(E)\mathbf{x} = b(E), \\
\text{minimize } \|A(L|E)\mathbf{x} - b(L|E)\|_\infty
\end{cases}
\end{align*}
$$

\tag{3}

with $A \in \mathcal{A}_{m,n}$, $E = \{1, \ldots, r\}$, $L = \{1, \ldots, m\}$, and $A(E) \in \mathcal{A}_{r,n}$ and the GMP :

$$
\begin{align*}
\begin{cases}
A(E)\mathbf{x} = b(E), \\
\text{minimize } \|A(I|E)\mathbf{x} - b(I|E)\|_\infty
\end{cases}
\end{align*}
$$

\tag{4}

with $A \in \mathcal{A}_{m,n}$, $E = \{1, \ldots, r\}$, $I \subseteq L$, and $A(E) \in \mathcal{A}_{r,n}$.

We shall denote the deviation of the GMP (3) by $\tau(A, b)$, and the deviation of the GMP (4) by $\tau(A(I), b(I))$.

The set of minimax solutions of GMP (3) will be denoted by $P(A, b)$ and the corresponding set of GMP (4) by $P(A(I), b(I))$. To (3) we shall refer as GMP, and to (4) as RGMP(restricted).

4.2.1. Theorem.

If the deviation of the GMP is $\rho(A, b)$ and the deviation of the OMP, minimize $\|Ax - b\|_\infty$, is $\rho(A, b)$, then $\rho(A, b) \leq \tau(A, b)$.

Proof. If $P(A, b) \subseteq T(A, b)$ then $\rho(A, b) = \tau(A, b)$, otherwise, $\rho(A, b) < \tau(A, b)$. □
4.2.2. Corollary.

Let \( K \in \mathcal{K}_k(A) \), then \( \rho(A(K), \mathfrak{h}(K)) \leq \tau(A, \mathfrak{h}). \)

\[ \square \]

4.2.3. Theorem.

Let \( K \in \mathcal{K}_k(A) \) and let \( \lambda \neq \emptyset \) so that \( \sum_{i \in K} \lambda_i a_i = \emptyset \) and for \( i \notin K \), \( \lambda_i = 0 \).

Consider the RGMP with \( I = K \).

Let

\[ \tau = (- \sum_{i \in K} \lambda_i b_i) \left( \sum_{i \in K \setminus E} \lambda_i \right)^{-1}, \]

then

\[ \tau(A(K), \mathfrak{h}(K)) \geq |\tau|. \]

**Proof.** Since \( A(E) \) has full rank, we conclude that \( A(K \setminus E) \) is non-trivial. Let \( y \in \mathbb{R}^n \) so that \( A(E)y = \mathfrak{h}(E) \), then

\[ \lambda^T x(y) = - \lambda^T b = \tau \left( \sum_{i \in K \setminus E} |\lambda_i| \right). \]

Hence

\[ |\tau| = \left| \sum_{i \in K \setminus E} \lambda_i r_i(x) \right| \left( \sum_{i \in K \setminus E} |\lambda_i| \right)^{-1} \leq \|x(y)\|_{\infty}, \quad (5) \]

and this completes the proof, since \( y \) was arbitrary.

\[ \square \]

**Remark.**

Contrary to theorem 2.2.1, only a lower-bound is found for the deviation in the special case that in (4) \( A(I) \) is a cadre matrix.

4.2.4. Corollary.

Let \( K \in \mathcal{K}_k(A) \) so that \( A(K) \) is the only cadre matrix of \( A(K \cup E) \).

Let \( \lambda \neq \emptyset \) so that \( \sum_{i \in K} \lambda_i a_i = \emptyset \) and for \( i \notin K \), \( \lambda_i = 0 \).

Let

\[ \tau = (- \sum_{i \in K} \lambda_i b_i) \left( \sum_{i \in K \setminus E} |\lambda_i| \right)^{-1}, \]

then
(i) \( \tau(A(K), b(K)) = |\tau| \),

(ii) for any \( x \in P(A(K), b(K)) \) holds that for \( i \in K \setminus E \),
\[ r_i(x) = \tau \text{ sign} (\lambda_i), \]

(iii) \( \tau(A(K), b(K)) \leq \tau(A, b) \).

Proof. \( \tau(A(K), b(K)) = |\tau| \) if and only if in (5) equality holds, or, equivalently, if and only if a \( y \in \mathbb{R}^n \) exists so that

\[
\begin{cases}
  i \in E, & a_i^T y = b_i, \\
  i \in K \setminus E, & a_i^T y = b_i + \tau \text{ sign} (\lambda_i).
\end{cases}
\]  

(6)

The rank of the matrix of coefficients of the system (6) is one less than the number of equations, since \( A(K \cup E) \) contains only one cadre matrix (see section 4).

Moreover, since
\[
\sum_{i \in E} \lambda_i a_i + \sum_{i \in K \setminus E} \lambda_i a_i = 0,
\]
and also
\[
\sum_{i \in E} \lambda_i b_i + \sum_{i \in K \setminus E} (\lambda_i b_i + \tau \lambda_i \text{ sign} (\lambda_i)) = 0,
\]
we conclude that indeed a \( y \in \mathbb{R}^n \) exists satisfying (6).

Next let us show that any GMP has the property that a cadre matrix \( A(K) \) exists so that \( \tau(A(K), b(K)) = \tau(A, b) \).

4.2.5. Theorem.

Consider the GMP.

There is a cadre matrix \( A(K) \) so that

(i) \( A(K \cup E) \) has only one cadre matrix,

(ii) \( \tau(A(K), b(K)) = \tau(A, b) \).

Proof. Let \( x \in P(A, b) \) be a minimax solution of the GMP.

Let \( I \) be defined by
\[ I := \{ i \mid |r_i(x)| = \tau(A, b) \}, \]
and let for \( i \in I \) scalars \( a_i \) be defined so that
\( \sigma_i = \text{sign} \left( r_i(x) \right) \).

Let for \( i \in I \), \( \mathbf{a}_i \) be written as

\[
\mathbf{a}_i = \mathbf{a}_i^1 + \mathbf{a}_i^\perp,
\]

where \( \mathbf{a}_i^1 \in S^1 \) and \( \mathbf{a}_i^\perp \in S := \text{Span of } \{ \mathbf{a}_i | i \in E \} \).

There is no \( \mathbf{u} \in \mathbb{R}^n \) so that

\[
\begin{cases}
  \text{for } i \in E, \mathbf{a}_i^1 \mathbf{u} = 0, \\
  \text{for } i \in I, \sigma_i \mathbf{a}_i^\perp \mathbf{u} > 0,
\end{cases}
\]

or, equivalently,

\[
\text{for } i \in I, (\sigma_i \mathbf{a}_i^1)^T \mathbf{u} > 0,
\]

since, otherwise, \( \mathbf{y} \in \mathbb{R}^n \) would exist with \( \| \mathbf{r}(\mathbf{y}) \|_\infty < \tau(A, b) \).

By lemma 1.1.2, the consequence is that

\[ \emptyset \in \text{CH}(\sigma_i \mathbf{a}_i^1 | i \in I). \]

By lemma 1.1.6, a set of indices \( I' \subset I \) exists so that the set

\( \{ \mathbf{a}_i^1 | i \in I' \} \) is a cadre matrix,

and

\[ \emptyset \in \text{CH}(\sigma_i \mathbf{a}_i^1 | i \in I'). \]

This implies that scalars \( \mu_i > 0 \) exist so that

\[
\sum_{i \in I'} \mu_i \sigma_i \mathbf{a}_i^1 = \emptyset.
\]

Let

\[ \mathbf{a}^* := \sum_{i \in I'} \mu_i \sigma_i \mathbf{a}_i^\perp. \]

\( \mathbf{a}^* \in S \), and, consequently, there are uniquely determined scalars \( \beta_i \) so that

\[ \mathbf{a}^* = -\sum_{i \in E} \beta_i \mathbf{a}_i. \]

Therefore,

\[
\sum_{i \in E} \beta_i \mathbf{a}_i + \sum_{i \in I'} \mu_i \sigma_i \mathbf{a}_i^1 = \emptyset,
\]
where the scalars $\beta_i$ and $\mu_i$ are uniquely determined up to multiplication by a scalar.

So $A(I' \cup E)$ contains only one cadre matrix and, thus, we obtain by theorem 4.2.4.

$$\tau(A(I'), b(I')) = \left| \sum_{i \in I} \nu_i \sigma_i \right| \tau(A, b) \left( \sum_{i \in I'} |\nu_i| \right)^{-1} = \tau(A, b),$$

and this proves the theorem.

**4.2.6. Corollary.**

There is a reference $A(I^*)$ so that

(i) $A(I^*) \supset A(E)$

(ii) $\tau(A(I^*), b(I)) = \tau(A, b)$.

**Proof.** It is possible to extend $I' \cup E$ to a reference, since the number of elements in $I' \cup E$ is less than or equal to $(n+1)$.

**Remark.**

In conformity with the previous sections we shall call a cadre matrix, as appearing in 4.2.5, a maximal cadre matrix, and a reference as in 4.2.6, a maximal reference.

**4.2.7. Definition.**

Considering the GMP, $x \in P(A, b)$ is called an extremal minimax solution if and only if

$$\text{rank } A(I|I = \{i \mid r_i(x) = \tau(A, b)\cup E) = n.}$$

In the foregoing we saw that the matrix of coefficients of a GMP contains a maximal cadre matrix $A(K)$ and a maximal reference $A(I)$ so that $A(I) \supset A(E)$. We found that, if $x$ is a minimax solution of the GMP, the components of the residual vector to $x$, corresponding with rows of $A(K|E)$ are independent of the special choice of $x$ in $P(A, b)$.

We also know that $P(A, b)$, being bounded and intersection of a finite number of slabs, is a polyhedron.

Without further proof we mention the theorems
4.2.8. Theorem.

$x \in P(A, b)$ is an extremal minimax solution of the GMP, if and only if $x$ is an extremal vector of $P(A, b)$, and

4.2.9. Theorem.

If $K$ is the only maximal cadre matrix of $A$ with respect to $b$ then the dimension of $P(A, b)$ equals $d$, with

$$d = n - k - \mu(E|K).$$

4.3. In this part we are concerned with the calculation of an extremal minimax solution of the GMP.

First of all the case will be considered that $A$ itself is a reference, then the theorem analogous to 3.3.1 will be stated and proved, and, finally, without further discussion, the generalization of the algorithm proposed in 3.5.1 is stated, which supplies an extremal minimax solution for the GMP.

4.3.1. Theorem.

Considering the GMP, let $A$ itself be a reference. Let $\lambda \neq \emptyset$ so that $\lambda^T A = \emptyset^T$.

Let $A(K)$ be the cadre matrix of $A$ and

$$\tau := (- \sum_{i \in K} \lambda_i b_i) (\sum_{i \in K \backslash E} |\lambda_i|)^{-1}.$$

Then,

(i) $\tau(A, b) = |\tau|$, 

(ii) if $x$ is an extremal vector of $P(A, b)$,

for $i \in E$, $r_i(x) = 0$, 

for $i \in K \backslash E$, $r_i(x) = \tau \text{ sign } (\lambda_i)$, 

for $i \in L \backslash (K \cup E)$, $|r_i(x)| = \tau(A, b)$,

(iii) $\dim P(A, b) = n - k - \mu(E|K)$. 


Proof.

(i) follows immediately from the fact that, if there is a maximal reference, it must be $A$ itself and, if there is a maximal cadre matrix, it must be $A(K)$.

(ii) and (iii) are obvious from 4.2. \qed

4.3.2. Definition (on the analogy of 3.1.1).

Let $A \in \mathcal{M}_{m,n}$, $b \in \mathbb{R}^m$, $E = \{1, \ldots, r\}$, $r < n$, and $K \in \mathcal{R}_k(A)$.

Let $\lambda \neq \emptyset$ and $\sum_{i \in K} \lambda_i a_i = \emptyset$, then

$$S_E(A(K), b(K)) := \{x | x \in \mathbb{R}^n; \exists \sigma, \forall i \in K : \text{sign} (r_i(x) \lambda_i) = \sigma\}. \qed$$

4.3.3. Theorem.

Considering the GMP, let $A \in \mathcal{M}_{n+2,n}$, $b \in \mathbb{R}^{n+2}$.

Let $I_1 \in \mathcal{R}_n(A)$ and $K_1 \in \mathcal{R}_k(A)$ so that $I_1 \supseteq K_1$ and $I_1 \supseteq E$.

Let $\tau(A(I_1), b(I_1)) > 0$.

Let $x_1$ be an extremal minimax solution of $P(A(I_1), b(I_1))$.

Let $\{\ell\} := L | I_1$, and $|r_\ell(x_1)| > \tau(A(I_1), b(I_1))$.

Then there is a (generally not uniquely determined) $I_2 \in \mathcal{R}_n(A)$, distinct from $I_1$, so that, if its cadre is $K_2$,

$$I_2 \supseteq E \text{ and } x_1 \in S_E(A(K_2), b(K_2)).$$

For such an $I_2$ holds

$$\tau(A(I_2), b(I_2)) \geq \tau(A(I_1), b(I_1)).$$

Proof. Let $\lambda \neq \emptyset$ so that $\sum_{i \in K_1} \lambda_i a_i = \emptyset$ and define for $i \in L | I_1$, $\lambda_i := 0$.

Without loss of generality we may assume

$$i \in K_1 | E, r_i(x_1) = \tau(A(I_1), b(I_1)) \text{ sign} (\lambda_i),$$

for if this is not so, then we may take $\lambda := -\lambda$.

Let $y$ be such that $\sum_{i \in I_1} v_i a_i = a_x$ and define $v_x := -1$.

We distinguish two cases.
1st. case, \(\forall i \in I_1 \mid (K_1 \cup E), \forall_i r_i(x_1) r_\ell(x_1) \leq 0\).

Choose \(s \in K_1 \mid E\) so that

\[
r_\ell(x_1) \frac{\nu_s}{\lambda_s} = \max_{i \in K_1 \mid E} \left( r_i(x_1) \frac{\nu_i}{\lambda_i} \right).
\]

Consider

\[ I_2 := (I_1 \mid \{s\}) \cup \{\ell\}. \]

It is obvious that \(I_2\) is a reference, that \(I_2\) contains \(E\), and that \(s\) needs not be unique.

Let \(K_2\) be the cadre of \(I_2\), then \(\ell \in K_2\).

Let \(\mu \neq \theta\) so that \(\sum_{i \in K_2} \mu_i a_i = \theta\), then \(\mu = \alpha \lambda - \beta \nu\) with \(\alpha \lambda_s - \beta \nu_s\) and \((\alpha, \beta) \neq (0, 0)\).

Since \(\lambda_s \neq 0\) we may choose \(\beta = 1\) and hence have

\[
\left\{ \begin{array}{ll}
\text{for } i \in K_2 \cap K_1, & \mu_i = \lambda_i \left( \frac{\nu_s}{\lambda_s} - \frac{\nu_i}{\lambda_i} \right), \\
\text{for } i \in K_2 \setminus K_1, & \mu_i = -\nu_i.
\end{array} \right.
\]

Using (*) and (**) we obtain

\[
\text{for } i \in K_2 \mid E, \text{ sign } (\mu_i r_i(x_1)) = \text{ sign } (r_\ell(x_1)),
\]

which implies that \(x_1 \in S_E(K_2)\).

Since

\[
\begin{align*}
\min_{i \in K_2 \mid E} |r_i(x_1)| &= \tau(A(I_1), b(I_1)) \\
\max_{i \in K_2 \mid E} |r_i(x_1)| &= |r_\ell(x_1)| > \tau(A(I_1), b(I_1)),
\end{align*}
\]

and for any \(x \in S_E(K_2)\) holds that

\[
\min_{i \in K_2 \mid E} |r_i(x)| \leq \tau(A(K_2), b(K_2)) \leq \max_{i \in K_2 \mid E} |r_i(x_1)|,
\]

with either twice equality or twice inequality,

we obtain

\[
\tau(A(I_2), b(I_2)) = \tau(A(K_2), b(K_2)) > \tau(A(I_1), b(I_1)).
\]
2nd. case, \( t \in I_1 \cap (K_1 \cup E), \) \( v_t r_t(x_1) r_\ell(x_1) > 0. \)
Consider
\[
I_2 := (I_1 \{s\}) \cup \{\ell\}.
\]
Since \( v_t \neq 0, \) \( I_2 \) must be a reference containing \( E. \)
\( I_1 \) and \( I_2 \) have the same cadre \( K_1, \) so \( x_1 \in S_E(K_1) = S_E(K_2), \) and, moreover,
\[
\tau(A(I_2), b(I_2)) = \tau(A(I_1), b(I_1)).
\]

4.3.4. Corollary.
\( T(A(I_2), b(I_2)) \) has an extremal vector \( x_2 \) so that
\[
\text{for } i \in I_2, \text{ sign } (r_i(x_2)) = \text{sign } (r_i(x_1)).
\]

4.3.5. Corollary.
There is an \( I_2, \) having a cadre \( K_2, \) satisfying
\[
\left\{ \begin{array}{l}
x_1 \in S_E(K_2), \\
\tau(A(I_2), b(I_2)) > \tau(A(I_1), b(I_1))
\end{array} \right.
\]
if and only if
\[
\forall i \in I_1 \, |K_1, \, v_i r_i(x_1) r_\ell(x_1) \leq 0.
\]

Remark.

If \( \tau(A(I_1), b(I_1)) = 0, \) then we choose \( s \) so that
\[
r_\ell(x_1) \frac{v_s}{\lambda_s} = \max_{i \in K_1 \, |E} \frac{r_i(x_1)}{\lambda_i}.
\]
With this choice of \( s \) it readily follows that \( \tau(A(I_2), b(I_2)) \neq 0. \)

We shall now formulate the analogon of the algorithm, proposed in 3.5.1., capable of finding an extremal minimax solution of the GMP.
4.3.6. The algorithm.

Let $A \in \mathbb{M}^{m,n}$, with $m > n$ and $b \in \mathbb{R}^m$.

Let $E = \{1, \ldots, r\}$ and $A(E) \in \mathbb{M}^{r,n}$.

Then

(0) $j := 0$,

(1) choose $I \in \mathcal{R}_n(A)$ so that $I \supset E$,

(2) determine $\lambda \neq \emptyset$ so that

$$\sum_{i \in I} \lambda_i a_i = \emptyset,$$

and compute

$$\tau(I) := (-\sum_{i \in I} \lambda_i b_i) \left( \sum_{i \in E} |\lambda_i| \right)^{-1},$$

after which, if necessary, $\lambda$ is adapted so that $\tau(I) \geq 0$,

(3) determine the cadre $K$ of $I$,

$$K := \{i| i \in I, \lambda_i \neq 0\},$$

and choose scalars $\sigma_i$ for $i \in I(K \cup E)$ so that $|\sigma_i| = 1$,

(4) determine

for $i \in E$, $r_i := 0$,

for $i \in K \cup E$, $r_i := \tau(I) \text{ sign}(\lambda_i)$,

for $i \in I(K \cup E)$, $r_i := \tau(I)\sigma_i$,

(5) compute the solution $x$ of

$$A(I)x = b(I) + r(I),$$

(6) determine $\ell \in L|I$ so that $|r_{\ell}(x)|$ is maximal, if $|r_{\ell}(x)| \leq \rho(I)$ then $x$ is a solution, otherwise

(7) compute $\nu$ so that

$$\delta_{\nu} = \sum_{i \in I} \nu_i \delta_i,$$

(8) if for some $s \in I(K \cup E)$ it appears that

$$\nu s r_{\ell}(x)u_i > 0,$$
then \( I := (I \setminus \{s\}) \cup \{l\}, j := j+1, \sigma_s := \text{sign } (r_\xi(x)) \) and return to (4), otherwise,

\[
(9) \quad \text{determine } M := \max_{i \in \mathcal{K} \setminus E} \left( r_\xi(x) \frac{v_i}{\lambda_i} \right) = r_\xi(x) \frac{v_s}{\lambda_s},
\]

\( I := (I \setminus \{s\}) \cup \{l\}, j = j+1, \) return to (2).

**Remark.**

It is obvious that any reference occurring during the sequence of iterations contains the set \( A(E) \).
References.

   Introduction to Approximation Theory, McGraw-Hill.

   Contribution au calcul des approximations de Tschebycheff, thesis.


   Convexity, Cambridge University Press.

   Linear Programming, McGraw-Hill.

   Convex Polytopes, John Wiley and Sons.

   Linear Algebra, Addison-Wesley Publishing Co.

   Linear Programming, Addison-Wesley Publishing Co.

   Convex Analysis, Princeton University Press.

    Convexity and Optimisation in finite dimensions, Springer.

    Über diskrete und lineare Tschebyscheff-Approximationen,

On the best linear Chebyshev approximations, Computer Journal, 10, 172-177.
