

On ℓ_p -approximate solutions of linear equations

Jianming Miao* and Adi Ben-Israel †

September 26, 1992
Revised: May 1, 1993

Abstract

For $1 \leq p \leq \infty$, the ℓ_p -**approximate solutions** of $A\mathbf{x} = \mathbf{b}$ are the minimizers of $\|A\mathbf{x} - \mathbf{b}\|_p$ where $\|\cdot\|_p$ is the ℓ_p -**norm**. We consider the special case where the null space of A^T is one-dimensional. Sample results:

(a) If $1 \leq p \leq \infty$ and A is $m \times (m-1)$ of rank $m-1$, then there is a matrix $A^{\{p\}}$ (depending on A and p) such that, for every $\mathbf{b} \in \mathbf{R}^m$, the vector $A^{\{p\}}\mathbf{b}$ is an ℓ_p -approximate solution of $A\mathbf{x} = \mathbf{b}$, which is unique if $1 < p < \infty$.

(b) If $1 < p < \infty$ and A is $m \times n$ of rank $m-1$, then there is a matrix $A_{\{2\}}^{\{p\}}$ (depending on A and p) such that for every $\mathbf{b} \in \mathbf{R}^m$ the vector $A_{\{2\}}^{\{p\}}\mathbf{b}$ is the ℓ_p -approximate solution of minimal euclidean norm.

(c) Let $1 \leq p \leq \infty$ and let A be $m \times n$ of rank $m-1$. Then there is a vector $\mathbf{r}^{\{p\}}$ (computed from any least squares solution of $A\mathbf{x} = \mathbf{b}$), such that any ordinary solution of the auxiliary equation

$$A\mathbf{x} = \mathbf{b} + \mathbf{r}^{\{p\}},$$

is an ℓ_p -approximate solution of $A\mathbf{x} = \mathbf{b}$.

Key words: Linear equations. ℓ_p -approximate solutions. Minimum ℓ_p -norm solutions. Least squares solutions. Generalized inverses. Moore-Penrose inverse.

1 Introduction

Given $A \in \mathbf{R}^{m \times n}$ and $\mathbf{b} \in \mathbf{R}^m$, consider the linear equation

$$A\mathbf{x} = \mathbf{b}, \tag{1.1}$$

*Supported by Grant NSF-STC91-19999.

†RUTCOR-Rutgers Center for Operations Research, Rutgers University, P.O. Box 5062, New Brunswick, NJ 08903-5062, U.S.A.

and for any $\mathbf{x} \in \mathbb{R}^n$ the **residual**

$$\mathbf{r} = \mathbf{r}(\mathbf{x}) := A\mathbf{x} - \mathbf{b}. \quad (1.2)$$

If (1.1) is inconsistent, we often must settle for a “solution” vector minimizing some norm of the residual. Using the ℓ_p -**norms**, defined for $1 \leq p \leq \infty$ and $\mathbf{u} = (u_j) \in \mathbb{R}^m$ by

$$\|\mathbf{u}\|_p := \begin{cases} \left(\sum_{j=1}^m |u_j|^p \right)^{\frac{1}{p}}, & 1 \leq p < \infty \\ \max_{1 \leq j \leq m} |u_j| & , p = \infty \end{cases}, \quad (1.3)$$

an ℓ_p -**approximate solution**, sometimes called a **best ℓ_p -approximate solution**, of (1.1) and denoted by $\mathbf{x}^{\{p\}}$, is a solution of the minimization problem

$$\min \{ \|A\mathbf{x} - \mathbf{b}\|_p : \mathbf{x} \in \mathbb{R}^n \}. \quad (1.4)$$

In particular, the ℓ_2 -approximate solutions are the **least squares solutions**, which have their advantages (nice analytical properties and efficient algorithms) as well as disadvantages (sensitivity to deviating observations, such as would result from experimental errors). Other ℓ_p -norms used in (1.4) include the ℓ_∞ -norm (giving **minimax** or **Chebyshev solutions**) [8] and the ℓ_1 -norm, see e.g. [5], [13]. However, for $p = 1$ and ∞ the objective function $\|A\mathbf{x} - \mathbf{b}\|_p$ is typically non-differentiable at the optimal solutions, making the latter hard to characterize and compute. For $p \neq 2$, the ℓ_p -approximate solutions are in general computed iteratively, see e.g. [12].

Some notation: The **set of increasing sequences** of r elements from $\{1, \dots, m\}$ is

$$Q_{r,m} := \{I = \{i_1, \dots, i_r\} : 1 \leq i_1 < i_2 < \dots < i_r \leq m\}$$

For $A \in \mathbb{R}_r^{m \times n}$, $r > 0$, we denote the following **index sets**:

$\mathcal{I}(A) := \{I \in Q_{r,m} : \text{rank } A_{I*} = r\}$, i.e. the **maximal sets of linearly independent rows**,

$\mathcal{J}(A) := \{J \in Q_{r,n} : \text{rank } A_{*J} = r\}$, i.e. the **maximal sets of linearly independent columns**,

$\mathcal{N}(A) := \{(I, J) \in Q_{r,m} \times Q_{r,n} : \text{rank } A_{IJ} = r\}$, i.e. the **maximal nonsingular submatrices**.

The index sets $\mathcal{I}(A)$, $\mathcal{J}(A)$ and $\mathcal{N}(A)$ shall be abbreviated here by \mathcal{I} , \mathcal{J} and \mathcal{N} respectively. We have

$$\mathcal{N} = \mathcal{I} \times \mathcal{J}, \quad \text{see e.g. [1]}. \quad (1.5)$$

The **basic solutions** of the linear equation $A\mathbf{x} = \mathbf{b}$ are the solutions of subsystems corresponding to maximal nonsingular submatrices of A . The basic solutions are for

$$A \text{ of full column-rank} : \{A_{I*}^{-1} \mathbf{b}_I : I \in \mathcal{I}\}, \quad (1.6)$$

$$\text{for } A \text{ of full row-rank} : \{\widehat{A_{*J}^{-1}} \mathbf{b} : J \in \mathcal{J}\}, \quad (1.7)$$

$$\text{and for general } A : \{\widehat{A_{IJ}^{-1}} \mathbf{b}_I : (I, J) \in \mathcal{N}\}, \quad (1.8)$$

where \mathbf{b}_I is the I^{th} subvector of \mathbf{b} , and $\widehat{}$ denotes a vector padded by zeros. The **convex hull of basic solutions** of the given equation $A\mathbf{x} = \mathbf{b}$ shall be denoted by $\mathcal{C} = \mathcal{C}(A, \mathbf{b})$.

Ben-Tal and Teboulle [3] showed that for A of full column-rank and $1 \leq p < \infty$, the ℓ_p -approximate solutions of (1.1) lie in $\mathcal{C} = \text{conv}\{A_{I^*}^{-1}\mathbf{b}_I : I \in \mathcal{I}\}$, the convex hull of the basic solutions (1.6). For $p = \infty$, there is an ℓ_∞ -approximate solution in \mathcal{C} . These results were extended in [10] to general matrices.

If there is more than one solution (or approximate solution), we often select one of minimal ℓ_p -norm, using not necessarily the same norm as in (1.4). In particular, the selection $p = 2$ is natural in statistical applications, where $\|\mathbf{x}\|_2$ is related to the variance of the estimate \mathbf{x} .

Given $1 \leq p_1, p_2 \leq \infty$, we define a **minimum ℓ_{p_2} -norm ℓ_{p_1} -approximate solution** of (1.1), denoted by $\mathbf{x}_{\{p_2\}}^{\{p_1\}}$ as a solution of the two-stage minimization problem :

$$\textbf{Stage 1} \quad \min_{\mathbf{x}} \{ \|A\mathbf{x} - \mathbf{b}\|_{p_1} : \mathbf{x} \in \mathbb{R}^n \} \quad (1.9)$$

$$\textbf{Stage 2} \quad \min_{\mathbf{x}} \{ \|\mathbf{x}\|_{p_2} : \text{among all solutions } \mathbf{x} \text{ of Stage 1} \} . \quad (1.10)$$

The solution set of (1.9) is a closed convex set. Therefore $\mathbf{x}_{\{p_2\}}^{\{p_1\}}$ is unique for $1 < p_2 < \infty$, since then the ℓ_p -norm is strictly convex.

The case $p_1 = p_2 = 2$ is the easiest to handle and the one most thoroughly studied. The **minimum ℓ_2 -norm least squares solution** of $A\mathbf{x} = \mathbf{b}$ is given in closed form as

$$\mathbf{x}_{\{2\}}^{\{2\}} = A^\dagger \mathbf{b} , \quad \forall \mathbf{b} \in \mathbb{R}^m , \quad (1.11)$$

where A^\dagger is the **Moore-Penrose inverse** of A . Moreover $\mathbf{x}_{\{2\}}^{\{2\}}$ is a convex combination of the basic solutions (1.8) (see e.g. [11], [4], [1])

$$\mathbf{x}_{\{2\}}^{\{2\}} = \sum_{(I,J) \in \mathcal{N}} \lambda_{IJ} \widehat{A_{IJ}^{-1}} \mathbf{b}_I , \quad (1.12)$$

where the weights λ_{IJ} are independent of \mathbf{b} ,

$$\lambda_{IJ} = \frac{\det^2 A_{IJ}}{\sum_{(K,L) \in \mathcal{N}} \det^2 A_{KL}} . \quad (1.13)$$

The corresponding result for the Moore-Penrose inverse of A is

$$A^\dagger = \sum_{(I,J) \in \mathcal{N}} \lambda_{IJ} \widehat{A_{IJ}^{-1}} , \quad (1.14)$$

where $\widehat{A_{IJ}^{-1}}$ is an $n \times m$ matrix which has the inverse of A_{IJ} in position (J, I) , and zeros elsewhere.

In this paper we consider the case of ℓ_p -norms, $1 \leq p \leq \infty$.

- Residuals, and approximation errors are studied in § 2.

- In § 3 we consider the special case where A is an $(n + 1) \times n$ matrix of rank n . The ℓ_p -approximate solutions $\mathbf{x}^{\{p\}}$ are given explicitly as convex combinations of basic solutions with convex weights which are independent of \mathbf{b} . In this case A has a generalized inverse $A^{\{p\}}$ in the sense that for every \mathbf{b} , $\mathbf{x}^{\{p\}} := A^{\{p\}} \mathbf{b}$ is an ℓ_p -approximate solution.
- § 4 deals with $A \in \mathbb{R}_{m-1}^{m \times n}$ and $1 < p < \infty$. There the minimum ℓ_2 -norm ℓ_p -approximate solutions $\mathbf{x}_{\{2\}}^{\{p\}}$ are convex combinations of basic solutions, with convex weights which are independent of \mathbf{b} . In this case there is a matrix $A_{\{2\}}^{\{p\}}$, which is a generalized inverse of A in the sense that for every $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{x}_{\{2\}}^{\{p\}} := A_{\{2\}}^{\{p\}} \mathbf{b}$ is a minimum ℓ_2 -norm ℓ_p -approximate solution.
- § 5 deals again with $A \in \mathbb{R}_{m-1}^{m \times n}$. The ℓ_p -residual $\mathbf{r}^{\{p\}} = \mathbf{r}(\mathbf{x}^{\{p\}})$ is computed in terms of the least squares residual $\mathbf{r}^{\{2\}}$, giving an auxiliary equation

$$A \mathbf{x} = \mathbf{b} + \mathbf{r}^{\{p\}}$$

whose solutions are the ℓ_p -approximate solutions of (1.1).

2 Residuals

If $\dim N(A) > 0$, the ℓ_p -approximate solutions of $A \mathbf{x} = \mathbf{b}$ are not unique. However, for any A and $1 < p < \infty$ the residual is unique:

Lemma 2.1 Given $1 < p < \infty$, the residual

$$\mathbf{r}^{\{p\}} := \mathbf{r}(\mathbf{x}^{\{p\}}) = A \mathbf{x}^{\{p\}} - \mathbf{b}$$

is the same for all ℓ_p -approximate solutions $\mathbf{x}^{\{p\}}$ of $A \mathbf{x} = \mathbf{b}$.

Proof: The uniqueness of $\mathbf{r}^{\{p\}}$ is guaranteed by the strict convexity of ℓ_p -norms, $1 < p < \infty$. \square

The following lemma describes the residuals in the special case of $A \in \mathbb{R}_{m-1}^{m \times n}$, which is the case when the range $R(A)$ of A is a hyperplane in \mathbb{R}^m , or equivalently, the null-space $N(A^T)$ of the transpose A^T is one-dimensional.

Lemma 2.2 Let $A \in \mathbb{R}_{m-1}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Then

- If $\mathbf{b} \notin R(A)$ then $N(A^T)$ is spanned by the least squares residual $\mathbf{r}^{\{2\}}$.
- The set of residuals

$$H := \{A \mathbf{x} - \mathbf{b} : \mathbf{x} \in \mathbb{R}^n\} \tag{2.1}$$

is a hyperplane of \mathbb{R}^m . In fact

$$H = \{\mathbf{r} : \langle \mathbf{r}, \mathbf{r}^{\{2\}} \rangle = \langle \mathbf{r}^{\{2\}}, \mathbf{r}^{\{2\}} \rangle\}. \tag{2.2}$$

Proof: (a) Clearly $N(A^T)$ is spanned by $\mathbf{r}^{\{2\}}$ since $\dim N(A^T) = 1$, $\mathbf{r}^{\{2\}} \neq \mathbf{0}$ and the normal equation

$$A^T (A\mathbf{x} - \mathbf{b}) = \mathbf{0} \quad (2.3)$$

characterizes least squares solutions.

(b) From $H = -\mathbf{b} + R(A)$, $\dim R(A) = m - 1$, and (a). \square

Remark 2.1 Part (b) of the lemma was proved in [8, Lemma, p. 40] under an additional (unnecessary) assumption that the columns of A are linearly independent.

The following theorem gives bounds for the error of ℓ_p -approximations. For $1 \leq p \leq \infty$, the **conjugate** $q = q(p)$ is defined as customary by

$$\frac{1}{p} + \frac{1}{q} = 1 \quad , \quad 1 < p < \infty \quad , \quad (2.4)$$

and the limits $q(1) := \lim_{p \rightarrow 1} q(p) = \infty$, $q(\infty) := \lim_{p \rightarrow \infty} q(p) = 1$.

Theorem 2.1 Let $A \in \mathbb{R}_r^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, $1 \leq p \leq \infty$ and q be the conjugate of p . Let $\mathbf{x}^{\{2\}}$ be any least squares solution of $A\mathbf{x} = \mathbf{b}$, and let

$$\mathbf{r}^{\{2\}} := A\mathbf{x}^{\{2\}} - \mathbf{b} \quad (2.5)$$

be the corresponding residual. Then

$$\frac{|\langle \mathbf{b}, \mathbf{r}^{\{2\}} \rangle|}{\|\mathbf{r}^{\{2\}}\|_q} \leq \min_{\mathbf{x} \in \mathbb{R}^n} \|A\mathbf{x} - \mathbf{b}\|_p \leq \|\mathbf{r}^{\{2\}}\|_p \quad (2.6)$$

Proof. The upper bound is trivial, since $\mathbf{r}^{\{2\}}$ is a residual, but not necessarily optimal. The lower bound is a consequence of the well-known duality theorem

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|A\mathbf{x} - \mathbf{b}\|_p = \max_{\mathbf{y} \in \mathbb{R}^m} \{\langle \mathbf{b}, \mathbf{y} \rangle : A^T \mathbf{y} = \mathbf{0} \text{ , } \|\mathbf{y}\|_q \leq 1\} \quad (2.7)$$

where $1 \leq p \leq \infty$ and q is the conjugate of p , see e.g. [6], [7]. Note that the least squares residual $\mathbf{r}^{\{2\}}$ is in $N(A^T)$ by (2.3). Therefore the left hand side of (2.6) is not greater than the right hand side of (2.7). \square

Remark 2.2 The bounds (2.6) are useful since a least squares residual is readily available, i.e.

$$\mathbf{r}^{\{2\}} = (I - AA^+) \mathbf{b} = P_{N(A^T)} \mathbf{b} \quad , \quad (2.8)$$

and (2.6) becomes

$$\frac{\|P_{N(A^T)} \mathbf{b}\|_2^2}{\|P_{N(A^T)} \mathbf{b}\|_q} \leq \min_{\mathbf{x} \in \mathbb{R}^n} \|A\mathbf{x} - \mathbf{b}\|_p \leq \|P_{N(A^T)} \mathbf{b}\|_p \quad (2.9)$$

Remark 2.3 If $A \in \mathbb{R}_{m-1}^{m \times n}$, it follows from Lemma 2.2(a) that the lower bounds in (2.6) and (2.9) are exact, namely

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|A\mathbf{x} - \mathbf{b}\|_p = \frac{|\langle \mathbf{b}, \mathbf{r}^{\{2\}} \rangle|}{\|\mathbf{r}^{\{2\}}\|_q} = \frac{\|P_{N(A^T)} \mathbf{b}\|_2^2}{\|P_{N(A^T)} \mathbf{b}\|_q}. \quad (2.10)$$

3 A is an $(n + 1) \times n$ matrix of rank n

Throughout this section let $A \in \mathbb{R}_n^{(n+1) \times n}$ and $\mathbf{b} \in \mathbb{R}^{n+1}$. The basic solutions of $A\mathbf{x} = \mathbf{b}$ are

$$\mathbf{x}_{I^*} := A_{I^*}^{-1} \mathbf{b}_I, \quad I \in \mathcal{I}, \quad (3.1)$$

and their convex hull, $\mathcal{C} := \text{conv}\{\mathbf{x}_{I^*} : I \in \mathcal{I}\}$. (3.2)

It follows from [3] that there exists a solution of (1.4) in \mathcal{C} . In this section we give explicit formulæ for the convex weights which turn out to be independent of \mathbf{b} . Therefore there is a matrix $A^{\{p\}}$ such that for any $\mathbf{b} \in \mathbb{R}^{n+1}$, $A^{\{p\}}\mathbf{b}$ is an ℓ_p -approximate solution. We shall construct the matrix $A^{\{p\}}$ and study its properties.

An immediate consequence of the assumption that A is $(n + 1) \times n$ of rank n is that for each $I \in \mathcal{I}$, the basic solution $\mathbf{x}_{I^*} = A_{I^*}^{-1} \mathbf{b}_I$ has a residual $\mathbf{r}_I = \mathbf{r}(\mathbf{x}_{I^*})$ with n zeros in positions I . The remaining component of \mathbf{r}_I , denoted by ϵ_I , shall now be given explicitly.

Lemma 3.1 Let $A \in \mathbb{R}_n^{(n+1) \times n}$. For any $I \in \mathcal{I}$ let

$$\epsilon_I := A_{i^*} \mathbf{x}_{I^*} - b_i, \quad (3.3)$$

where $\{i\}$ is the complement of I in $\{1, \dots, n + 1\}$, A_{i^*} is the i^{th} row of A , and b_i is the i^{th} component of \mathbf{b} . Then

$$|\epsilon_I| = \frac{|\det(A, \mathbf{b})|}{|\det A_{I^*}|}. \quad (3.4)$$

Proof. This follows from

$$\det \begin{pmatrix} A_{I^*} & \mathbf{b}_I \\ A_{i^*} & b_i \end{pmatrix} = (b_i - A_{i^*} A_{I^*}^{-1} \mathbf{b}_I) \det A_{I^*}. \quad \square$$

Another consequence of assuming $A \in \mathbb{R}_n^{(n+1) \times n}$ is that for any vector \mathbf{x} which is a convex combination of the basic solutions,

$$\mathbf{x} = \sum_{I \in \mathcal{I}} \mu_I \mathbf{x}_{I^*}, \quad \sum_{I \in \mathcal{I}} \mu_I = 1, \quad \mu_I \geq 0, \quad (3.5)$$

the residual $\mathbf{r}(\mathbf{x})$ has components $(\mu_I \epsilon_I)$. Therefore the problem of minimizing the ℓ_p -norm of residuals of the vectors in (3.5) can be written as

$$(P) \quad \min \left\{ \|(\mu_I |\epsilon_I|)\|_p : \sum_{I \in \mathcal{I}} \mu_I = 1, \mu_I \geq 0 \right\},$$

with $|\epsilon_I|$ given by (3.4), $I \in \mathcal{I}$. This problem can be solved explicitly for all $1 \leq p \leq \infty$: First the case $p = 1$.

Theorem 3.1 Let $A \in \mathbf{R}_n^{(n+1) \times n}$. Then

$$\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|_1 = \frac{|\det(A, \mathbf{b})|}{|\det A_{L^*}|}, \quad (3.6)$$

and an optimal solution is $\mathbf{x}^{\{1\}} = A_{L^*}^{-1} \mathbf{b}_L$, (3.7)

where

$$L \in \arg \max_I |\det A_{I^*}|, \quad (3.8)$$

and $\arg \max f$ [$\arg \min f$] denote the **set of maximizers** [**minimizers**] of a function f .

Proof. By [3], any ℓ_1 -approximate solution is given by

$$\mathbf{x}^{\{1\}} = \sum_{I \in \mathcal{I}} \mu_I \mathbf{x}_{I^*},$$

where

$$\sum_{I \in \mathcal{I}} \mu_I = 1, \quad \mu_I \geq 0.$$

Then

$$\begin{aligned} \|A\mathbf{x}^{\{1\}} - \mathbf{b}\|_1 &= \sum_{I \in \mathcal{I}} \mu_I |\epsilon_I|, \\ &= |\det(A, \mathbf{b})| \sum_{I \in \mathcal{I}} \frac{\mu_I}{|\det A_{I^*}|} \quad \text{by (3.4)}. \end{aligned}$$

For $p = 1$ the optimization problem (P) can therefore be written as

$$\min \left\{ \sum_{I \in \mathcal{I}} \frac{\mu_I}{|\det A_{I^*}|} : \text{s.t. } \sum_{I \in \mathcal{I}} \mu_I = 1, \mu_I \geq 0 \right\},$$

and it has an optimal solution

$$\begin{cases} \mu_L^{(1)} = 1 & \text{for } L \text{ selected in } \arg \max_I |\det A_{I^*}|, \\ \mu_I^{(1)} = 0 & \text{if } I \neq L. \end{cases} \quad \square$$

Remark 3.1 The optimal basis L is independent of \mathbf{b} and the solution is unique if and only if $|\det A_{L^*}| > |\det A_{I^*}|$ for all $I \neq L$, $I \in \mathcal{I}$.

We now solve for $1 < p < \infty$. In this case the solution is unique.

Theorem 3.2 Let $A \in \mathbf{R}_n^{(n+1) \times n}$, $1 < p < \infty$, and q be the conjugate (2.4) of p . Then

$$\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|_p = \frac{|\det(A, \mathbf{b})|}{\left(\sum_{K \in \mathcal{I}} |\det A_{K*}|^q\right)^{\frac{1}{q}}}, \quad (3.9)$$

and the optimal solution is

$$\mathbf{x}^{\{p\}} = \sum_{I \in \mathcal{I}} \mu_I^{(p)} \mathbf{x}_{I*}, \quad (3.10)$$

where

$$\mu_I^{(p)} = \frac{|\det A_{I*}|^q}{\sum_{K \in \mathcal{I}} |\det A_{K*}|^q}, \quad I \in \mathcal{I}. \quad (3.11)$$

Proof. The solution is given by

$$\mathbf{x}^{\{p\}} = \sum_{I \in \mathcal{I}} \mu_I \mathbf{x}_{I*},$$

where

$$\sum_{I \in \mathcal{I}} \mu_I = 1, \quad \mu_I \geq 0.$$

Then

$$\begin{aligned} \|A\mathbf{x}^{\{p\}} - \mathbf{b}\|_p^p &= \sum_{I \in \mathcal{I}} \mu_I^p |\epsilon_I|^p, \\ &= |\det(A, \mathbf{b})|^p \sum_{I \in \mathcal{I}} \frac{\mu_I^p}{|\det A_{I*}|^p}. \end{aligned}$$

The optimization problem (P) here becomes

$$\min \left\{ \sum_{I \in \mathcal{I}} \frac{\mu_I^p}{|\det A_{I*}|^p} : \sum_{I \in \mathcal{I}} \mu_I = 1, \mu_I \geq 0 \right\}. \quad (3.12)$$

The Kuhn-Tucker necessary and sufficient conditions for (3.12) are

$$\begin{aligned} \frac{\mu_I^{p-1}}{|\det A_{I*}|^p} &= c, \quad \forall I \in \mathcal{I}, \\ \sum_{I \in \mathcal{I}} \mu_I &= 1, \\ \mu_I &\geq 0, \end{aligned}$$

for some constant c . The results (3.9) and (3.10)–(3.11) then follow immediately. \square

Remark 3.2 The weights $\mu_I^{(p)}$ are independent of \mathbf{b} .

Finally the case $p = \infty$:

Theorem 3.3 Let $A \in \mathbf{R}_n^{(n+1) \times n}$. Then

$$\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|_{\infty} = \frac{|\det(A, \mathbf{b})|}{\sum_{K \in \mathcal{I}} |\det A_{K*}|}, \quad (3.13)$$

and an optimal solution is

$$\mathbf{x}^{\{\infty\}} = \sum_{I \in \mathcal{I}} \mu_I^{(\infty)} \mathbf{x}_{I*}, \quad (3.14)$$

where

$$\mu_I^{(\infty)} = \frac{|\det A_{I*}|}{\sum_{K \in \mathcal{I}} |\det A_{K*}|}, \quad I \in \mathcal{I}. \quad (3.15)$$

Proof. By [3], there is a solution $\mathbf{x}^{\{\infty\}}$ in \mathcal{C} . Let

$$\mathbf{x}^{\{\infty\}} = \sum_{I \in \mathcal{I}} \mu_I \mathbf{x}_{I*},$$

where

$$\sum_{I \in \mathcal{I}} \mu_I = 1, \quad \mu_I \geq 0.$$

Then

$$\begin{aligned} \|A\mathbf{x}^{\{\infty\}} - \mathbf{b}\|_{\infty} &= \max_{I \in \mathcal{I}} \mu_I |\epsilon_I|, \\ &= |\det(A, \mathbf{b})| \max_{I \in \mathcal{I}} \frac{\mu_I}{|\det A_{I*}|}. \end{aligned}$$

The optimization problem (P) here becomes

$$\min \left\{ \max_{I \in \mathcal{I}} \frac{\mu_I}{|\det A_{I*}|} : \sum_{I \in \mathcal{I}} \mu_I = 1, \mu_I \geq 0 \right\},$$

with

$$\mu_I^{(\infty)} = \frac{|\det A_{I*}|}{\sum_{K \in \mathcal{I}} |\det A_{K*}|}, \quad I \in \mathcal{I},$$

as optimal solution. □

Remark 3.3 For $p = \infty$ the weights $\mu_I^{(\infty)}$ are again independent of \mathbf{b} . It follows, e.g. [8, p. 42, Problems 6,7], that the solution is unique if and only if the **Haar condition** is satisfied, i.e., if every set of n rows of A is linearly independent.

Definition 3.1 For any $A \in \mathbf{R}_n^{(n+1) \times n}$ define the matrix $A^{\{p\}}$

$$A^{\{p\}} := \begin{cases} \widehat{A}_{L*}^{-1} & \text{if } p = 1, \text{ and } L \text{ is selected in } \arg \max_I |\det A_{I*}|, \\ \sum_{I \in \mathcal{I}} \mu_I^{(p)} \widehat{A}_{I*}^{-1} & \text{if } 1 < p \leq \infty, \end{cases} \quad (3.16)$$

where \widehat{A}_{I*}^{-1} is an $n \times (n+1)$ matrix with the inverse of A_{I*} in position I , and zeros elsewhere.

Example 3.1 Let $A = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$. Then

$$A^{\{p\}} = \begin{cases} \left(0, \frac{1}{2} \right) & \text{if } p = 1, \\ \frac{1}{1+2^q} (1, 2^{q-1}) & \text{for all } 1 < p \leq \infty, \end{cases}$$

where q is the conjugate of p .

Theorems 3.1–3.3 can be summarized as follows:

Corollary 3.1 Let $A \in \mathbb{R}_n^{(n+1) \times n}$, and let $A^{\{p\}}$ be given by (3.16). Then for any $\mathbf{b} \in \mathbb{R}^{n+1}$, $A^{\{p\}} \mathbf{b}$ is a solution of (1.4). \square

The following example shows that the conclusion does not hold for $A \in \mathbb{R}_k^{(n+1) \times k}$ with $k < n$.

Example 3.2 Let $A = \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix}$, $\mathbf{b}_1 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$, $\mathbf{b}_2 = \begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix}$, $\mathbf{b}_3 = \mathbf{b}_1 + \mathbf{b}_2 = \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix}$.

The corresponding solutions of $\min \|A\mathbf{x} - \mathbf{b}_i\|_4$ are

$$\mathbf{x}_1 = -0.321465, \quad \mathbf{x}_2 = 0.334298, \quad \mathbf{x}_3 = 0.053283,$$

respectively. However

$$\mathbf{x}_3 \neq \mathbf{x}_1 + \mathbf{x}_2.$$

The matrix $A^{\{p\}}$ defined in (3.16) is a generalized inverse of A . Clearly $A^{\{p\}}$ is a left-inverse,

$$A^{\{p\}} A = I, \tag{3.17}$$

and consequently $A^{\{p\}}$ satisfies

$$AXA = A, \quad XAX = X, \quad (XA)^T = XA, \tag{3.18}$$

showing that $A^{\{p\}}$ is a **{1,2,4}-inverse** of A , see [2].

Theorem 3.4 Let $A \in \mathbb{R}_n^{(n+1) \times n}$ and $Q \in \mathbb{R}_n^{n \times n}$. Then

$$(AQ)^{\{p\}} = Q^{-1} A^{\{p\}}. \tag{3.19}$$

Proof. Let $B = AQ$, then $B_{I^*} = A_{I^*} Q$, and $\mathcal{I}(B) = \mathcal{I}(A) = \mathcal{I}$. Therefore

$$\begin{aligned} B^{\{p\}} &= \sum_{I \in \mathcal{I}} \mu_I^{(p)} \widehat{B_{I^*}^{-1}}, \\ &= Q^{-1} \sum_{I \in \mathcal{I}} \mu_I^{(p)} \widehat{A_{I^*}^{-1}}, \end{aligned}$$

where $\mu_I^{(p)}$ as above. □

Now consider the continuity of $A^{\{p\}}$: If $\{A_j\} \subset \mathbb{R}^{(n+1) \times n}$ is a sequence of matrices converging to a matrix $A \in \mathbb{R}_n^{(n+1) \times n}$, then eventually all $\{A_j\}$ have full column-rank n . Does the corresponding sequence of generalized inverses $A_j^{\{p\}}$ converge to $A^{\{p\}}$? The affirmative answer given below, for $1 < p < \infty$, shows $A^{\{p\}}$ to be insensitive to errors in A .

Theorem 3.5 Let $1 < p < \infty$ and let $A, A_j \in \mathbb{R}_n^{(n+1) \times n}$ be such that $A_j \rightarrow A$. Then

$$A_j^{\{p\}} \rightarrow A^{\{p\}} \quad \text{as } j \rightarrow \infty. \quad (3.20)$$

Proof. Since $A_j \rightarrow A$, there is an index j_0 such that for any $j \geq j_0$

$$\mathcal{I}(A) \subseteq \mathcal{I}(A_j).$$

Since

$$A_j^{\{p\}} = \sum_{I \in \mathcal{I}(A_j)} (\mu_j)_I^{(p)} (\widehat{A_j})_{I^*}^{-1},$$

we need only consider terms $(\mu_j)_I^{(p)} (\widehat{A_j})_{I^*}^{-1}$ for $I \in \mathcal{I}(A_j) \setminus \mathcal{I}(A)$. For such I it follows from $A_j \rightarrow A$ that

$$\det(A_j)_{I^*} \rightarrow 0,$$

and

$$(\mu_j)_I^{(p)} (\widehat{A_j})_{I^*}^{-1} = \frac{|\det(A_j)_{I^*}|^{(q-1)}}{\sum_{K \in \mathcal{I}(A_j)} |\det(A_j)_{K^*}|^q} \text{adj}((A_j)_{I^*}) \rightarrow 0,$$

where $\text{adj}((A_j)_{I^*})$ is the **adjoint** of $(A_j)_{I^*}$. This completes the proof. □

Remark 3.4 If the set $\arg \max_I |\det A_{I^*}|$ is a singleton, then $A^{\{1\}}$ is continuous by the continuity of $A_{L^*}^{-1}$. Otherwise $A^{\{1\}}$ depends on the choice of L .

Remark 3.5 If the Haar condition is satisfied, then $A^{\{\infty\}}$ is continuous by the proof of Theorem 3.5. The next example shows that in general $A^{\{\infty\}}$ need not be continuous.

Example 3.3 Let $A = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $A_\epsilon = \begin{pmatrix} 1 \\ \epsilon \end{pmatrix}$. Then

$$A^{\{\infty\}} = (1, 0), \quad A_\epsilon^{\{\infty\}} = \left(\frac{1}{1+\epsilon}, \frac{1}{1+\epsilon} \right)$$

$$\text{and } A_\epsilon^{\{\infty\}} \not\rightarrow A^{\{\infty\}}, \quad \text{as } \epsilon \rightarrow 0.$$

For $1 \leq p \leq \infty$, let $\mathbf{r}^{\{p\}}$ be the residual of $\mathbf{x}^{\{p\}}$:

$$\mathbf{r}^{\{p\}} = A\mathbf{x}^{\{p\}} - \mathbf{b}. \quad (3.21)$$

Then we have

Theorem 3.6 All the residuals $\mathbf{r}^{\{p\}}$ have the same sign vectors for $1 < p \leq \infty$ (or may be so chosen in the event of nonuniqueness).

Proof. Let $\mathbf{x}^{\{p\}}$ be the solution of (1.4) as given by Theorems 3.2–3.3, and let $\{i\}$ be the complement of I in $\{1, \dots, n+1\}$. Then

$$r_i^{\{p\}} = \begin{cases} \mu_I^{(p)} \epsilon_I & \text{if } I \in \mathcal{I}, \\ 0 & \text{if } I \notin \mathcal{I}. \end{cases}$$

The result then follows from Theorems 3.2–3.3. \square

The next example shows that Theorem 3.6 does not hold for $p = 1$, notwithstanding an assertion in [8, p. 40] which would imply that all the residuals have the same sign vectors for $1 \leq p \leq \infty$ (or may be so chosen in the event of nonuniqueness).

Example 3.4 Let $A = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$. Then $\mathbf{x}^{\{1\}} = \frac{1}{2}$, $\mathbf{x}^{\{2\}} = 1$, and their residuals

$$\mathbf{r}^{\{1\}} = \begin{pmatrix} -\frac{5}{2} \\ 0 \end{pmatrix}, \quad \mathbf{r}^{\{2\}} = \begin{pmatrix} -2 \\ -1 \end{pmatrix},$$

do not agree in sign.

4 Minimum ℓ_2 -norm ℓ_p -approximate solutions

For $A \in \mathbf{R}_{m-1}^{m \times n}$ and $1 < p < \infty$ we can give explicit formulæ for the solutions of

$$\min_{\mathbf{x}} \{ \|\mathbf{x}\|_2 : \mathbf{x} \in \arg \min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|_p \}. \quad (4.1)$$

The results require additional hypotheses for $p = \infty$ and $p = 1$, see Theorem 4.1(b)–(c).

For any full-rank factorization $A = CR$, the problem (4.1) can be solved in stages

$$\min_{\mathbf{y}} \|C\mathbf{y} - \mathbf{b}\|_p, \quad (4.2)$$

$$\min_{\mathbf{x}} \{ \|\mathbf{x}\|_2 : R\mathbf{x} = \mathbf{y}, \mathbf{y} \in \arg \min_{\mathbf{y}} \|C\mathbf{y} - \mathbf{b}\|_p \}, \quad (4.3)$$

using the basic solutions (1.8)

$$\mathbf{x}_{IJ} := A_{IJ}^{-1} \widehat{\mathbf{b}}_I, \quad (I, J) \in \mathcal{N}. \quad (4.4)$$

Theorem 4.1 Let $A \in \mathbf{R}_{m-1}^{m \times n}$ and $\mathbf{b} \in \mathbf{R}^m$.

(a) Let $1 < p < \infty$. Then the solution of (4.1) is given by

$$\mathbf{x}_{\{2\}}^{\{p\}} = \sum_{I \in \mathcal{I}} \mu_I A_{I*}^\dagger \mathbf{b}_I, \quad (4.5)$$

$$= \sum_{(I,J) \in \mathcal{N}} \lambda_{IJ} \mathbf{x}_{IJ}. \quad (4.6)$$

The weights λ_{IJ} are

$$\lambda_{IJ} := \mu_I \nu_J, \quad (4.7)$$

and

$$\mu_I = \frac{\text{vol}^q A_{I*}}{\sum_{K \in \mathcal{I}} \text{vol}^q A_{K*}}, \quad \nu_J = \frac{\text{vol}^2 A_{*J}}{\sum_{N \in \mathcal{J}} \text{vol}^2 A_{*N}}, \quad (4.8)$$

where for a matrix $X \in \mathbf{R}_m^{m \times n}$ the **volume** is $\text{vol } X := \sqrt{\det XX^T}$, see [1].

(b) Let $p = \infty$. If every set of $m - 1$ rows of A is linearly independent, then the solution $\mathbf{x}_{\{2\}}^{\{\infty\}}$ is given by (4.5)–(4.8) with $q = 1$.

(c) Let $p = 1$. If there is $L \in \mathcal{I}$ such that for some $J \in \mathcal{J}$

$$|\det A_{LJ}| > |\det A_{IJ}|, \quad \forall I \neq L, I \in \mathcal{I}, \quad (4.9)$$

then the solution $\mathbf{x}_{\{2\}}^{\{1\}}$ of

$$\min_{\mathbf{x}} \{\|\mathbf{x}\|_2 : \mathbf{x} \in \arg \min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_1\} \quad (4.10)$$

is

$$\mathbf{x}_{\{2\}}^{\{1\}} = A_{L*}^\dagger \mathbf{b}_L. \quad (4.11)$$

Proof. Let $A = CR$ be any full-rank factorization of A . Then for any $I \in \mathcal{I}$

$$A_{I*} = C_{I*}R, \quad A_{I*}^\dagger = R^\dagger C_{I*}^{-1}, \quad (4.12)$$

and

$$\text{vol } A_{I*} = |\det C_{I*}| \text{vol } R. \quad (4.13)$$

(a) The solution of (4.2) is given from (Theorem 3.2) by

$$\mathbf{y} = \sum_{I \in \mathcal{I}} \mu_I C_{I*}^{-1} \mathbf{b}_I, \quad (4.14)$$

where

$$\mu_I = \frac{|\det C_{I*}|^q}{\sum_{K \in \mathcal{I}} |\det C_{K*}|^q} = \frac{\text{vol}^q A_{I*}}{\sum_{K \in \mathcal{I}} \text{vol}^q A_{K*}} \quad \text{by (4.13)}.$$

Then the solution of (4.3) is given by

$$\begin{aligned} \mathbf{x}_{\{2\}}^{\{p\}} &= R^\dagger \mathbf{y}, \\ &= \sum_{I \in \mathcal{I}} \mu_I A_{I*}^\dagger \mathbf{b}_I, \quad \text{from (4.14), (4.12)}. \end{aligned} \quad (4.15)$$

(4.6) follows from (4.15) and

$$R^\dagger = \sum_{J \in \mathcal{J}} \nu_J \widehat{R_{*J}^{-1}}, \text{ see [1],}$$

where

$$\nu_J = \frac{|\det R_{*J}|^2}{\sum_{N \in \mathcal{J}} |\det R_{*N}|^2} = \frac{\text{vol}^2 A_{*J}}{\sum_{N \in \mathcal{J}} \text{vol}^2 A_{*N}}.$$

(b) The hypothesis is equivalent to a Haar condition on the full-column factor C . It follows from Remark 3.3 that the solution of $\min \|C\mathbf{y} - \mathbf{b}\|_\infty$ is unique. The proof then follows from Theorem 3.3.

(c) By hypothesis we have

$$|\det C_{L*}| > |\det C_{I*}|, \quad \forall I \neq L, \quad I \in \mathcal{I}.$$

Then Remark 3.1 implies the uniqueness of the solution of $\min \|C\mathbf{y} - \mathbf{b}\|_1$. The proof follows then from Theorem 3.1, similarly to the proof of part (a). \square

If the assumption in Theorem 4.1(c) is not satisfied, the convex weights may depend on \mathbf{b} even if the solution of (4.10) still lies in \mathcal{C} .

Example 4.1 Let $A = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$ where $b_1, b_2 > 0$. Then the basic solutions are $\mathbf{x}_1 = b_1$, $\mathbf{x}_2 = -b_2$, and $\arg \min \|A\mathbf{x} - \mathbf{b}\|_1 = \text{conv}\{b_1, -b_2\}$. Clearly the solution of (4.10) is

$$\mathbf{x}_{\{2\}}^{\{1\}} = \mathbf{0} = \lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2$$

with weights $\lambda_1 = \frac{b_2}{b_1 + b_2}$, $\lambda_2 = \frac{b_1}{b_1 + b_2}$ which depend on \mathbf{b} .

Definition 4.1 For any $A \in \mathbb{R}_{m-1}^{m \times n}$, $1 < p < \infty$, let the matrix $A_{\{2\}}^{\{p\}}$ be defined as

$$A_{\{2\}}^{\{p\}} = \sum_{(I,J) \in \mathcal{N}} \lambda_{IJ} \widehat{A_{IJ}^{-1}}, \quad (4.16)$$

with weights λ_{IJ} given by (4.8).

If $A \in \mathbb{R}_n^{(n+1) \times n}$ then Definition 4.1 reduces to Definition 3.1. From Theorem 4.1 we have

Corollary 4.1 Let $A \in \mathbb{R}_{m-1}^{m \times n}$, $1 < p < \infty$. Then for any $\mathbf{b} \in \mathbb{R}^m$, $A_{\{2\}}^{\{p\}} \mathbf{b}$ gives the solution of (4.1).

From Theorem 4.1 we can also deduce:

Theorem 4.2 Let $A \in \mathbf{R}_{m-1}^{m \times n}$ and $1 < p < \infty$. Then for any full-rank factorization $A = CR$

$$A_{\{2\}}^{\{p\}} = R^\dagger C^{\{p\}}, \quad (4.17)$$

where $C^{\{p\}}$ is defined by (3.16). Moreover $A_{\{2\}}^{\{p\}}$ is a $\{1, 2, 4\}$ -inverse of A .

Proof. (4.17) follows directly from the proof of Theorem 4.1. It is easy to check that $A_{\{2\}}^{\{p\}}$ satisfies the three equations in (3.18). \square

Theorem 4.3 Let $1 < p < \infty$ and let $A, A_j \in \mathbf{R}_{m-1}^{m \times n}$ be such that $A_j \rightarrow A$. Then

$$(A_j)_{\{2\}}^{\{p\}} \rightarrow A_{\{2\}}^{\{p\}} \quad \text{as } j \rightarrow \infty. \quad (4.18)$$

Proof. This follows by Theorem 3.5 and Theorem 4.2. \square

5 Auxiliary linear equations for ℓ_p -approximate solutions

In the special case that $A \in \mathbf{R}_{m-1}^{m \times n}$ we can construct an auxiliary linear equation whose (ordinary) solutions are the ℓ_p -approximate solutions of $A\mathbf{x} = \mathbf{b}$. For this we define the sign of a vector $\mathbf{z} = (z_j)$ as the vector

$$\text{sign } \mathbf{z} := (\text{sign } z_j). \quad (5.1)$$

The **Hadamard product** $\mathbf{u} \circ \mathbf{v}$ of two vectors $\mathbf{u} = (u_j)$ and $\mathbf{v} = (v_j)$ is the vector

$$\mathbf{u} \circ \mathbf{v} := (u_j v_j). \quad (5.2)$$

Theorem 5.1 Let $A \in \mathbf{R}_{m-1}^{m \times n}$, $\mathbf{b} \in \mathbf{R}^m$, $1 < p < \infty$, and q be the conjugate of p . Then the residual $\mathbf{r}^{\{p\}} := A\mathbf{x}^{\{p\}} - \mathbf{b}$ is given as

$$\mathbf{r}^{\{p\}} = \beta (\text{sign } \mathbf{r}^{\{2\}}) \circ |\mathbf{r}^{\{2\}}|^{q-1}, \quad (5.3)$$

where

$$\beta = \frac{\sum_{i=1}^m |r_i^{\{2\}}|^2}{\sum_{i=1}^m |r_i^{\{2\}}|^q}. \quad (5.4)$$

Proof: If $\mathbf{b} \in R(A)$ then $\mathbf{r}^{\{p\}} = \mathbf{0}$ for all p and (5.3) holds trivially for all β . We can therefore assume that $\mathbf{b} \notin R(A)$. From

$$\nabla \|A\mathbf{x} - \mathbf{b}\|_p^p = 0 \iff A^T \left((\text{sign } \mathbf{r}^{\{p\}}) \circ |\mathbf{r}^{\{p\}}|^{p-1} \right) = \mathbf{0} \quad (5.5)$$

and Lemma 2.2(a) it follows that

$$(\text{sign } \mathbf{r}^{\{p\}}) \circ |\mathbf{r}^{\{p\}}|^{p-1} = \alpha \mathbf{r}^{\{2\}},$$

for some scalar α which, written componentwise, becomes

$$\begin{aligned} (\text{sign } r_i^{\{p\}}) |r_i^{\{p\}}|^{p-1} &= \alpha r_i^{\{2\}}, \quad i = 1, \dots, m, \\ \text{so that } \text{sign } r_i^{\{p\}} &= \text{sign } \alpha \text{sign } r_i^{\{2\}}, \quad i = 1, \dots, m. \end{aligned}$$

Therefore

$$\begin{aligned} |r_i^{\{p\}}| &= |\alpha|^{q-1} |r_i^{\{2\}}|^{q-1}, \\ \text{and } r_i^{\{p\}} &= \beta (\text{sign } r_i^{\{2\}}) |r_i^{\{2\}}|^{q-1} \end{aligned}$$

for $\beta = (\text{sign } \alpha) |\alpha|^{q-1}$. We can determine the scalar β from the identity

$$\langle \mathbf{r}^{\{p\}}, \mathbf{r}^{\{2\}} \rangle = \langle \mathbf{r}^{\{2\}}, \mathbf{r}^{\{2\}} \rangle \quad (5.6)$$

which holds for all p by Lemma 2.2(b). Using (5.3) and (5.6) we obtain

$$\beta = \frac{\langle \mathbf{r}^{\{2\}}, \mathbf{r}^{\{2\}} \rangle}{\langle |\mathbf{r}^{\{2\}}|^{q-1}, |\mathbf{r}^{\{2\}}| \rangle} = \frac{\sum_{i=1}^m |r_i^{\{2\}}|^2}{\sum_{i=1}^m |r_i^{\{2\}}|^q}.$$

□

For $p = 1$ and $p = \infty$, the residuals $\mathbf{r}^{\{p\}}$ are no longer unique, but among them there exist residuals which can be given in terms of $\mathbf{r}^{\{2\}}$ analogously to (5.3). First the case $p = \infty$:

Theorem 5.2 Let $A \in \mathbf{R}_{m-1}^{m \times n}$, $\mathbf{b} \in \mathbf{R}^m$. Then there is a residual $\mathbf{r}^{\{\infty\}}$ satisfying

$$\mathbf{r}^{\{\infty\}} = \beta \text{sign } \mathbf{r}^{\{2\}}, \quad (5.7)$$

where

$$\beta = \frac{\sum_{i=1}^m |r_i^{\{2\}}|^2}{\sum_{i=1}^m |r_i^{\{2\}}|}. \quad (5.8)$$

Proof. This follows from Lemma 2.2(b) and [8, Problem 1, p. 42]. □

Theorem 5.3 Let $A \in \mathbf{R}_{m-1}^{m \times n}$, $\mathbf{b} \in \mathbf{R}^m$. Then there is a residual $\mathbf{r}^{\{1\}}$ satisfying

$$\mathbf{r}^{\{1\}} = \frac{\sum_{i=1}^m |r_i^{\{2\}}|^2}{r_\ell^{\{2\}}} \mathbf{e}_\ell, \quad (5.9)$$

where ℓ is selected in $\arg \max_i |r_i^{\{2\}}|$, and \mathbf{e}_ℓ denotes the ℓ^{th} unit vector.

Proof. This follows from Lemma 2.2(b) and [8, Problem 3, p. 42]. \square

From Theorems 5.1–5.3 it follows that the ℓ_p -approximate solutions can be computed as ordinary solutions of auxiliary linear equations, which are given in terms of the least-squares residual. We summarize these results as follows.

Corollary 5.1 Let $A \in \mathbf{R}_{m-1}^{m \times n}$.

(a) Let $1 < p < \infty$. Then the ℓ_p -approximate solutions of $A\mathbf{x} = \mathbf{b}$ are exactly the solutions of the linear equation

$$A\mathbf{x} = \mathbf{b} + \beta (\text{sign } \mathbf{r}^{\{2\}}) \circ |\mathbf{r}^{\{2\}}|^{q-1}, \quad (5.10)$$

where β is given by (5.4).

(b) Let $p = \infty$. Then any solution of (5.10), with β given by (5.8) and $q = 1$, is an ℓ_∞ -approximate solution of $A\mathbf{x} = \mathbf{b}$.

(c) Let $p = 1$ and let ℓ be selected in $\arg \max_i |r_i^{\{2\}}|$. Then any solution of the linear equation

$$A\mathbf{x} = \mathbf{b} + \frac{\sum_{i=1}^m |r_i^{\{2\}}|^2}{r_\ell^{\{2\}}} \mathbf{e}_\ell \quad (5.11)$$

is an ℓ_1 -approximate solution of $A\mathbf{x} = \mathbf{b}$. \square

Remark 5.1 For $A \in \mathbf{R}_n^{(n+1) \times n}$ and $p = \infty$, Corollary 5.1(b) reduces to [8, Theorem, p. 41], see also [9], [14].

Acknowledgement. We thank the referees for constructive suggestions.

References

- [1] A. Ben-Israel, “A volume associated with $m \times n$ matrices”, *Lin. Alg. Appl.* **167** (1992), 87-111.
- [2] A. Ben-Israel and T.N.E. Greville, *Generalized Inverses: Theory and Applications*, Wiley-Interscience, 1974.
- [3] A. Ben-Tal and M. Teboulle, “A geometric property of the least squares solution of linear equations”, *Lin. Alg. Appl.* **139** (1990), 165-170.
- [4] L. Berg, “Three results in connection with inverse matrices”, *Lin. Alg. Appl.* **84** (1986), 63-77.

- [5] P. Bloomfield and W.L. Steiger, *Least Absolute Deviations : Theory, Applications, and Algorithms*, Birkhäuser, Boston, 1983.
- [6] R.C. Buck, “Applications of duality in approximation theory”, pp. 27-42 in *Proceedings of Symposium on Approximation of Functions* (H.L. Garabedian, Ed.), Elsevier, New York, 1965.
- [7] P.L. Butzer and K. Scherer, “On fundamental theorems of approximation theory and their dual versions”, *J. Approx. Th.* **3** (1970), 87-100.
- [8] E.W. Cheney, *Introduction to Approximation Theory*, McGraw-Hill, New York, 1966.
- [9] M. Meicler, “Chebyshev solution of an inconsistent system of $n + 1$ linear equations in n unknowns in terms of its least squares solution”, *SIAM Review* **10** (1968), 373-375.
- [10] J. Miao and A. Ben-Israel, “The geometry of basic, approximate and minimum norm solutions of linear equations”, *Lin. Alg. Appl.* (to appear)
- [11] E.H. Moore, “On the reciprocal of the general algebraic matrix”, (Abstract), *Bull. Amer. Math. Soc.* **26** (1920), 394-395.
- [12] H. Späth, *Mathematical Algorithms for Linear Regression*, Academic Press, Boston, 1991.
- [13] V. Sposito, W. Smith and G. McCormick, *Minimizing the Sum of Absolute Deviations*, Vandenhoeck & Ruprecht, 1978.
- [14] R.P. Tewarson, “Minimax solution of $n + 1$ inconsistent linear equations in n unknowns”, *Computing* **5** (1970), 371-376.