On $\ell_p$–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 $\ell_p$–approximate solutions of $Ax = b$ are the minimizers of $\|Ax - b\|_p$ where $\| \cdot \|_p$ is the $\ell_p$–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 $b \in \mathbb{R}^m$, the vector $A^{(p)}b$ is an $\ell_p$–approximate solution of $Ax = b$, 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 $A^{(p)}_{(2)}$ (depending on $A$ and $p$) such that for every $b \in \mathbb{R}^m$ the vector $A^{(p)}_{(2)}b$ is the $\ell_p$–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 $r^{(p)}$ (computed from any least squares solution of $Ax = b$), such that any ordinary solution of the auxiliary equation

$$Ax = b + r^{(p)},$$

is an $\ell_p$–approximate solution of $Ax = b$.


1 Introduction

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

$$Ax = b, \quad (1.1)$$

*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 \( x \in \mathbb{R}^n \) the **residual**

\[
    r = r(x) := Ax - b.
\]

(1.2)

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

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

(1.3)

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

\[
    \min \left\{ \|Ax - b\|_p : x \in \mathbb{R}^n \right\}.
\]

(1.4)

In particular, the **\( \ell_2 \)-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 \( \ell_p \)-norms used in (1.4) include the **\( \ell_\infty \)-norm** (giving minimax or Chebyshev solutions) [8] and the **\( \ell_1 \)-norm**, see e.g. [5], [13]. However, for \( p = 1 \) and \( \infty \) the objective function \( \|Ax - b\|_p \) is typically non-differentiable at the optimal solutions, making the latter hard to characterize and compute. For \( p \neq 2 \), the **\( \ell_p \)-approximate solutions** are in general computed iteratively, see e.g. [12].

Some notation: The **set of increasing sequences** of \( r \) elements from \( \{1, \ldots, m\} \) is

\[
    Q_{r,m} := \{ I = \{i_1, \ldots, i_r\} : 1 \leq i_1 < i_2 < \cdots < i_r \leq m \}
\]

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

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

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

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

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

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

The **basic solutions** of the linear equation \( Ax = b \) are the solutions of subsystems corresponding to maximal nonsingular submatrices of \( A \). The basic solutions are for

\[
    A \text{ of full column-rank} : \{ A_{I*}^{-1} b_I : I \in \mathcal{I} \}, \tag{1.6}
\]

for \( A \) of full row-rank : \( \{ A_{*J}^{-1} b_J : J \in \mathcal{J} \}, \tag{1.7} \)

and for general \( A \) : \( \{ A_{IJ}^{-1} b_I : (I, J) \in \mathcal{N} \} \), \tag{1.8}
where \( b_I \) is the \( I^{th} \) subvector of \( b \), and \( \hat{} \) denotes a vector padded by zeros. The **convex hull of basic solutions** of the given equation \( A x = b \) shall be denoted by \( C = C(A, b) \).

Ben-Tal and Teboulle [3] showed that for \( A \) of full column-rank and \( 1 \leq p < \infty \), the \( \ell_p \)-approximate solutions of (1.1) lie in \( C = \text{conv}\{ A_I^{-1} b_I : I \in \mathcal{I} \} \), the convex hull of the basic solutions (1.6). For \( p = \infty \), there is an \( \ell_\infty \)-approximate solution in \( C \). 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 \( \ell_p \)-norm, using not necessarily the same norm as in (1.4). In particular, the selection \( p = 2 \) is natural in statistical applications, where \( \|x\|_2 \) is related to the variance of the estimate \( x \).

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

**Stage 1**

\[
\min_x \{ \| A x - b \|_{p_1} : x \in \mathbb{R}^n \} \tag{1.9}
\]

**Stage 2**

\[
\min_x \{ \| x \|_{p_2} : \text{among all solutions } x \text{ of Stage 1} \} \tag{1.10}
\]

The solution set of (1.9) is a closed convex set. Therefore \( x_{\{p_1\} \{p_2\}} \) is unique for \( 1 < p_2 < \infty \), since then the \( \ell_p \)-norm is strictly convex.

The case \( p_1 = p_2 = 2 \) is the easiest to handle and the one most thoroughly studied. The **minimum \( \ell_2 \)-norm least squares solution** of \( A x = b \) is given in closed form as

\[
x_{\{2\} \{2\}} = A^\dagger b , \quad \forall b \in \mathbb{R}^m , \tag{1.11}
\]

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

\[
x_{\{2\} \{2\}} = \sum_{(I,J) \in \mathcal{N}} \lambda_{IJ} \hat{A}_{IJ}^{-1} b_I , \tag{1.12}
\]

where the weights \( \lambda_{IJ} \) are independent of \( b \),

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

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

\[
A^\dagger = \sum_{(I,J) \in \mathcal{N}} \lambda_{IJ} \hat{A}_{IJ}^{-1} , \tag{1.14}
\]

where \( \hat{A}_{IJ}^{-1} \) 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 \( \ell_p \)-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 $\ell_p$–approximate solutions $x^{(p)}$ are given explicitly as convex combinations of basic solutions with convex weights which are independent of $b$. In this case $A$ has a generalized inverse $A^{(p)}$ in the sense that for every $b$, $x^{(p)} := A^{(p)} b$ is an $\ell_p$–approximate solution.

• § 4 deals with $A \in \mathbb{R}^{m \times n}$ and $1 < p < \infty$. There the minimum $\ell_2$–norm $\ell_p$–approximate solutions $x^{(p)}_{(2)}$ are convex combinations of basic solutions, with convex weights which are independent of $b$. In this case there is a matrix $A^{(p)}_{(2)}$, which is a generalized inverse of $A$ in the sense that for every $b \in \mathbb{R}^m$, $x^{(p)}_{(2)} := A^{(p)}_{(2)} b$ is a minimum $\ell_2$–norm $\ell_p$–approximate solution.

• § 5 deals again with $A \in \mathbb{R}^{m \times n}$. The $\ell_p$–residual $r^{(p)} = r\left(x^{(p)}\right)$ is computed in terms of the least squares residual $r^{(2)}$, giving an auxiliary equation

$$Ax = b + r^{(p)}$$

whose solutions are the $\ell_p$–approximate solutions of (1.1).

2 Residuals

If dim $N(A) > 0$, the $\ell_p$–approximate solutions of $Ax = b$ are not unique. However, for any $A$ and $1 < p < \infty$ the residual is unique:

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

$$r^{(p)} := r\left(x^{(p)}\right) = Ax^{(p)} - b$$

is the same for all $\ell_p$–approximate solutions $x^{(p)}$ of $Ax = b$.

**Proof:** The uniqueness of $r^{(p)}$ is guaranteed by the strict convexity of $\ell_p$–norms, $1 < p < \infty$. \hfill $\square$

The following lemma describes the residuals in the special case of $A \in \mathbb{R}^{m \times n}_{m-1}$, 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 $A \in \mathbb{R}^{m \times n}_{m-1}$ and $b \in \mathbb{R}^m$. Then

(a) If $b \not\in R(A)$ then $N(A^T)$ is spanned by the least squares residual $r^{(2)}$.

(b) The set of residuals

$$H := \{Ax - b : x \in \mathbb{R}^n\}$$

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

$$H = \{r : <r, r^{(2)}> = <r^{(2)}, r^{(2)}>\} .$$
Proof: (a) Clearly $N(A^T)$ is spanned by $r^{(2)}$ since $\dim N(A^T) = 1$, $r^{(2)} \neq 0$ and the normal equation
\[ A^T (Ax - b) = 0 \quad (2.3) \]
characterizes least squares solutions.
(b) From $H = -b + R(A)$, $\dim R(A) = m - 1$, 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 $\ell_p$-approximations. For $1 \leq p \leq \infty$, the conjugate $q = q(p)$ is defined as customary by
\[ \frac{1}{p} + \frac{1}{q} = 1 , \quad 1 < p < \infty , \quad (2.4) \]
and the limits
\[ q(1) := \lim_{p \to 1} q(p) = \infty , \quad q(\infty) := \lim_{p \to \infty} q(p) = 1 . \]

Theorem 2.1 Let $A \in \mathbb{R}_r^{m \times n}$, $b \in \mathbb{R}^m$, $1 \leq p \leq \infty$ and $q$ be the conjugate of $p$. Let $x^{(2)}$ be any least squares solution of $Ax = b$, and let
\[ r^{(2)} := Ax^{(2)} - b \quad (2.5) \]
be the corresponding residual. Then
\[ \left| \frac{\langle b, r^{(2)} \rangle}{\|r^{(2)}\|_q} \right| \leq \min_{x \in \mathbb{R}^n} \|Ax - b\|_p \leq \|r^{(2)}\|_p . \quad (2.6) \]

Proof. The upper bound is trivial, since $r^{(2)}$ is a residual, but not necessarily optimal. The lower bound is a consequence of the well-known duality theorem
\[ \min_{x \in \mathbb{R}^n} \|Ax - b\|_p = \max_{y \in \mathbb{R}^m} \{ \langle b, y \rangle : A^T y = 0 , \|y\|_q \leq 1 \} \quad (2.7) \]
where $1 \leq p \leq \infty$ and $q$ is the conjugate of $p$, see e.g. [6], [7]. Note that the least squares residual $r^{(2)}$ 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.
\[ r^{(2)} = (I - AA^+)b = P_{N(A^T)}b , \quad (2.8) \]
and (2.6) becomes
\[ \frac{\|P_{N(A^T)}b\|_2^2}{\|P_{N(A^T)}b\|_q} \leq \min_{x \in \mathbb{R}^n} \|Ax - b\|_p \leq \|P_{N(A^T)}b\|_p . \quad (2.9) \]
Remark 2.3 If $A \in \mathbb{R}^{m \times n}$, it follows from Lemma 2.2(a) that the lower bounds in (2.6) and (2.9) are exact, namely
\[
\min_{x \in \mathbb{R}^n} \|Ax - b\|_p = \frac{|\langle b, r^{(2)} \rangle |}{\|r^{(2)}\|_q} = \frac{\|P_N(A^T)b\|_2^2}{\|P_N(A^T)b\|_q}.
\] (2.10)

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

Throughout this section let $A \in \mathbb{R}^{(n+1) \times n}$ and $b \in \mathbb{R}^{n+1}$. The basic solutions of $Ax = b$ are
\[
x_{I^*} := A_{I^*}^{-1}b_I, \quad I \in \mathcal{I},
\] (3.1)
and their convex hull,
\[
C := \text{conv}\{x_{I^*} : I \in \mathcal{I}\}.
\] (3.2)

It follows from [3] that there exists a solution of (1.4) in $C$. In this section we give explicit formulæ for the convex weights which turn out to be independent of $b$. Therefore there is a matrix $A^{(p)}$ such that for any $b \in \mathbb{R}^{n+1}$, $A^{(p)}b$ is an $\ell_p$-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 $x_{I^*} = A_{I^*}^{-1}b_I$ has a residual $r_I = r(x_{I^*})$ with $n$ zeros in positions $I$. The remaining component of $r_I$, denoted by $\epsilon_I$, shall now be given explicitly.

Lemma 3.1 Let $A \in \mathbb{R}^{(n+1) \times n}$. For any $I \in \mathcal{I}$ let
\[
\epsilon_I := A_{i^*}^{-1}b_I - b_i,
\] (3.3)
where $\{i\}$ is the complement of $I$ in $\{1, \ldots, n+1\}$, $A_{i^*}$ is the $i^{th}$ row of $A$, and $b_i$ is the $i^{th}$ component of $b$. Then
\[
|\epsilon_I| = \frac{|\det (A, b)|}{|\det A_{i^*}|}.
\] (3.4)

Proof. This follows from
\[
\det \begin{pmatrix} A_{i^*} & b_I \\ A_{i^*} & b_i \end{pmatrix} = (b_i - A_{i^*}A_{i^*}^{-1}b_I) \det A_{i^*}.
\] □

Another consequence of assuming $A \in \mathbb{R}^{(n+1) \times n}$ is that for any vector $x$ which is a convex combination of the basic solutions,
\[
x = \sum_{I \in \mathcal{I}} \mu_I x_{I^*}, \quad \sum_{I \in \mathcal{I}} \mu_I = 1, \quad \mu_I \geq 0,
\] (3.5)
the residual $r(x)$ has components $(\mu_I \epsilon_I)$. Therefore the problem of minimizing the $\ell_p$-norm of residuals of the vectors in (3.5) can be written as
\[
(P) \quad \min \left\{ \|\mu_I \epsilon_I\|_p : \sum_{I \in \mathcal{I}} \mu_I = 1, \quad \mu_I \geq 0 \right\},
\]
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 $A \in \mathbb{R}^{(n+1)\times n}$. Then

$$
\min_x \|Ax - b\|_1 = \frac{|\det(A, b)|}{|\det A_L|},
$$

and an optimal solution is

$$
x^{(1)} = A_L^{-1}b_L,
$$

where

$$
L \in \arg \max_I |\det A_I|,
$$

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

**Proof.** By [3], any $\ell_1$–approximate solution is given by

$$
x^{(1)} = \sum_{I \in \mathcal{I}} \mu_I x_{I^*},
$$

where

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

Then

$$
\|Ax^{(1)} - b\|_1 = \sum_{I \in \mathcal{I}} \mu_I |\epsilon_I|,
$$

$$
= |\det(A, b)| \sum_{I \in \mathcal{I}} \frac{\mu_I}{|\det A_I|} \text{ by (3.4)}.
$$

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

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

and it has an optimal solution

$$
\left\{ \begin{array}{l}
\mu_L^{(1)} = 1 \quad \text{for } L \text{ selected in } \arg \max_I |\det A_I|, \\
\mu_I^{(1)} = 0 \quad \text{if } I \neq L.
\end{array} \right.
$$

**Remark 3.1** The optimal basis $L$ is independent of $b$ 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 $A \in \mathbb{R}^{(n+1) \times n}$, $1 < p < \infty$, and $q$ be the conjugate (2.4) of $p$. Then
\[
\min_{x} \|Ax - b\|_p = \frac{|\det(A, b)|}{\left(\sum_{K \in \mathcal{I}} |\det A_K|^q\right)^{\frac{1}{q}}},
\tag{3.9}
\]
and the optimal solution is
\[
x^{(p)} = \sum_{I \in \mathcal{I}} \mu_I^{(p)} x_{I^*},
\tag{3.10}
\]
where
\[
\mu_I^{(p)} = \frac{|\det A_I|^q}{\sum_{K \in \mathcal{I}} |\det A_K|^q}, \quad I \in \mathcal{I}.
\tag{3.11}
\]

**Proof.** The solution is given by
\[
x^{(p)} = \sum_{I \in \mathcal{I}} \mu_I x_{I^*},
\]
where
\[
\sum_{I \in \mathcal{I}} \mu_I = 1, \quad \mu_I \geq 0.
\]
Then
\[
\|Ax^{(p)} - b\|_p^p = \sum_{I \in \mathcal{I}} \mu_I^p |\epsilon_I|^p,
\]
\[
= |\det(A, b)|^p \sum_{I \in \mathcal{I}} \frac{\mu_I^p}{|\det A_I|^p}.
\]
The optimization problem (P) here becomes
\[
\min \left\{ \sum_{I \in \mathcal{I}} \frac{\mu_I^p}{|\det A_I|^p} : \sum_{I \in \mathcal{I}} \mu_I = 1, \mu_I \geq 0 \right\}.
\tag{3.12}
\]
The Kuhn-Tucker necessary and sufficient conditions for (3.12) are
\[
\frac{\mu_I^{p-1}}{|\det A_I|^p} = c, \quad \forall I \in \mathcal{I},
\]
\[
\sum_{I \in \mathcal{I}} \mu_I = 1, \quad \mu_I \geq 0,
\]
for some constant $c$. The results (3.9) and (3.10)–(3.11) then follow immediately. \qed

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

Finally the case $p = \infty$: 8
Theorem 3.3 Let $A \in \mathbb{R}^{(n+1) \times n}$. Then

$$
\min_x \|Ax - b\|_\infty = \frac{|\det(A, b)|}{\sum_{K \in \mathcal{I}} |\det A_{K^*}|},
$$

(3.13)

and an optimal solution is

$$
x^{(\infty)} = \sum_{I \in \mathcal{I}} \mu^{(\infty)}_I x_{I^*},
$$

(3.14)

where

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

(3.15)

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

$$
x^{(\infty)} = \sum_{I \in \mathcal{I}} \mu_I x_{I^*},
$$

where

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

Then

$$
\|Ax^{(\infty)} - b\|_\infty = \max_{I \in \mathcal{I}} \mu_I |\epsilon_I| = |\det(A, b)| \max_{I \in \mathcal{I}} \frac{\mu_I}{|\det A_{I^*}|}.
$$

The optimization problem (P) here becomes

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

with

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

as optimal solution. \qed

**Remark 3.3** For $p = \infty$ the weights $\mu^{(\infty)}_I$ are again independent of $b$. 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 $A \in \mathbb{R}^{(n+1) \times n}$ define the matrix $A^{(p)}$

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

(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
\[
A^{(p)} = \begin{cases} \begin{pmatrix} 0 & \frac{1}{2} \end{pmatrix} & \text{if } p = 1, \\ \frac{1}{1 + 2^p} \begin{pmatrix} 1 & 2^{q-1} \end{pmatrix} & \text{for all } 1 < p \leq \infty, \end{cases}
\]
where \( q \) is the conjugate of \( p \).

Theorems 3.1–3.3 can be summarized as follows:

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

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

Example 3.2 Let \( A = \begin{pmatrix} 1 & -2 \\ -1 & 3 \end{pmatrix} \), \( b_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \), \( b_2 = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \), \( b_3 = b_1 + b_2 = \begin{pmatrix} 0 \\ 2 \end{pmatrix} \).

The corresponding solutions of \( \min \|Ax - b_i\|_4 \) are
\[
x_1 = -0.321465, \ x_2 = 0.334298, \ x_3 = 0.053283,
\]
respectively. However
\[
x_3 \neq x_1 + x_2.
\]
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
\[
AXA = A, \ XAX = X, \ (XA)^T =XA, \tag{3.18}
\]
showing that \( A^{(p)} \) is a \( \{1,2,4\} \)-inverse of \( A \), see [2].

Theorem 3.4 Let \( A \in \mathbb{R}^{(n+1)\times n} \) and \( Q \in \mathbb{R}^{n\times n} \). Then
\[
(AQ)^{(p)} = Q^{-1}A^{(p)} . \tag{3.19}
\]

Proof. Let \( B = AQ \), then \( B_{I_*} = A_{I_*}Q \), and \( I(B) = I(A) = I \). Therefore
\[
B^{(p)} = \sum_{I \in \mathcal{I}} \mu_I^{(p)} B_{I_*}^{-1},
\]
\[
= Q^{-1} \sum_{I \in \mathcal{I}} \mu_I^{(p)} A_{I_*}^{-1},
\]
respectively.
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 $A \in \mathbb{R}^{n \times n}$, then eventually all $\{A_j\}$ have full column-rank $n$. Does the corresponding sequence of generalized inverses $A^{(p)}_j$ converge to $A^{(p)}$? The affirmative answer given below, for $1 < p < \infty$, shows $A^{(p)}_j$ to be insensitive to errors in $A$.

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

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

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

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

Since

$$A^{(p)}_j = \sum_{I \in \mathcal{I}(A_j)} (\mu_j)_I^{(p)} (A_j)^{-1}_I,$$

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

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

and

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

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

**Remark 3.4** If the set $\arg \max_I |\det A_{I^*}|$ is a singleton, then $A^{(1)}$ is continuous by the continuity of $A^{-1}_L$. 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

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

and $A_\epsilon^{(\infty)} \not\to A^{(\infty)}$, as $\epsilon \to 0$. \hfill 11
For $1 \leq p \leq \infty$, let $r^{(p)}$ be the residual of $x^{(p)}$:

$$r^{(p)} = Ax^{(p)} - b.$$  \hfill (3.21)

Then we have

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

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

$$r^{(p)}_{\{i\}} = \begin{cases} 
\mu^{(p)}_I \epsilon_I & \text{if } I \in \mathcal{I}, \\
0 & \text{if } I \notin \mathcal{I}.
\end{cases}$$

The result then follows from Theorems 3.2–3.3. \hfill \Box

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 $A = \begin{pmatrix} 1 & -2 \\ -2 & 5 \end{pmatrix}$, $b = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$. Then $x^{(1)} = \frac{1}{2}$, $x^{(2)} = 1$, and their residuals

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

do not agree in sign.

### 4 Minimum $\ell_2$–norm $\ell_p$–approximate solutions

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

$$\min_{x} \{\|x\|_2 : x \in \arg \min_{x} \|Ax - b\|_p\}.$$  \hfill (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

$$\min_{y} \|Cy - b\|_p,$$  \hfill (4.2)

$$\min_{x} \{\|x\|_2 : Rx = y, y \in \arg \min_{y} \|Cy - b\|_p\}.$$  \hfill (4.3)

using the basic solutions (1.8)

$$x_{IJ} := A_{IJ}^{-1} b_I, \quad (I, J) \in \mathcal{N}.$$  \hfill (4.4)
**Theorem 4.1** Let \( A \in \mathbb{R}^{m \times n} \) and \( b \in \mathbb{R}^m \).

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

\[
x^{(p)}_{(2)} = \sum_{I \in \mathcal{I}} \mu_I A^\dagger_I b_I ,
\]

where

\[
The weights \lambda_{IJ} are \lambda_{IJ} := \mu_I \nu_J ,
\]

and

\[
\mu_I = \frac{\text{vol}^q A_{I*}}{\sum_{K \in \mathcal{I}} \text{vol}^q A_{K*}} , \quad \nu_J = \frac{\text{vol}^2 A_{sJ}}{\sum_{N \in \mathcal{J}} \text{vol}^2 A_{sN}} ,
\]

(b) Let \( p = \infty \). If every set of \( m - 1 \) rows of \( A \) is linearly independent, then the solution \( x^{(\infty)}_{(2)} \) 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 \( J \in \mathcal{J} \)

\[
| \det A_{LJ} | > | \det A_{IJ} | , \quad \forall I \neq L , \quad I \in \mathcal{I} ,
\]

then the solution \( x^{(1)}_{(2)} \) of

\[
\min_{x} \{ \| x \|_2 : x \in \arg \min_{x} \| Ax - b \|_1 \}
\]

is

\[
x^{(1)}_{(2)} = A^\dagger_L b_L .
\]

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

\[
A_{I*} = C_{I*} R , \quad A^\dagger_{I*} = R^t C^{-1}_{I*} ,
\]

and

\[
\text{vol} A_{I*} = | \det C_{I*} | \text{vol} R .
\]

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

\[
y = \sum_{I \in \mathcal{I}} \mu_I C^{-1}_{I*} b_I ,
\]

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

\[
x^{(p)}_{(2)} = R^t y ,
\]

\[
= \sum_{I \in \mathcal{I}} \mu_I A^\dagger_I b_I , \quad \text{from (4.14), (4.12)}.
\]
(4.6) follows from (4.15) and

\[ R^\dagger = \sum_{J \in \mathcal{J}} \nu_J \widetilde{R}_s^\dagger, \quad \text{see [1]}, \]

where

\[ \nu_J = \frac{|\det R_s|^2}{\sum_{N \in \mathcal{J}} |\det R_s|^2} = \frac{\text{vol}^2 A_{sJ}}{\sum_{N \in \mathcal{J}} \text{vol}^2 A_{sN}}. \]

(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 \( \min \|Cy - b\|_\infty \) is unique. The proof then follows from Theorem 3.3.

(c) By hypothesis we have

\[ |\det C_L| > |\det C_I|, \quad \forall I \neq L, I \in \mathcal{I}. \]

Then Remark 3.1 implies the uniqueness of the solution of \( \min \|Cy - b\|_1 \). The proof follows then from Theorem 3.1, similarly to the proof of part (a).

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

**Example 4.1** Let \( A = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix} \) and \( b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \) where \( b_1, b_2 > 0 \). Then the basic solutions are \( x_1 = b_1, x_2 = -b_2 \), and \( \arg \min \|Ax - b\|_1 = \text{conv} \{b_1, -b_2\} \). Clearly the solution of (4.10) is

\[ x^{\{1\}}_{\{2\}} = 0 = \lambda_1 x_1 + \lambda_2 x_2 \]

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

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

\[ A^{\{p\}}_{\{2\}} = \sum_{(I,J) \in \mathcal{N}} \lambda_{IJ} \widetilde{A}_{IJ}^\dagger, \quad (4.16) \]

with weights \( \lambda_{IJ} \) given by (4.8).

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

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

From Theorem 4.1 we can also deduce:
**Theorem 4.2** Let $A \in \mathbb{R}^{m \times n}_{m-1}$ and $1 < p < \infty$. Then for any full-rank factorization $A = CR$

$$A_{(2)}^{(p)} = R^t C^{(p)} ,$$

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 $A, A_j \in \mathbb{R}^{m \times n}_{m-1}$ be such that $A_j \rightarrow A$. Then

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

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

### 5 Auxiliary linear equations for $\ell_p$–approximate solutions

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

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

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

$$u \circ v := (u_jv_j) .$$

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

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

where

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

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

$$\nabla \|Ax - b\|_p^p = 0 \iff A^T (\text{sign } r^{(p)}) \circ |r^{(p)}|^{p-1} = 0$$

15
and Lemma 2.2(a) it follows that
\[(\text{sign } r^{(p)}) \circ |r^{(p)}|^{p-1} = \alpha r^{(2)},\]
for some scalar \(\alpha\) which, written componentwise, becomes
\[(\text{sign } r_i^{(p)}) |r_i^{(p)}|^{p-1} = \alpha r_i^{(2)}, \quad i = 1, \ldots, m,
so that \(\text{sign } r_i^{(p)} = \text{sign } \alpha \text{ sign } r_i^{(2)}, \quad i = 1, \ldots, m.\)
Therefore
\[|r_i^{(p)}| = |\alpha|^{q-1} |r_i^{(2)}|^{q-1},\]
and \(r_i^{(p)} = \beta (\text{sign } r_i^{(2)}) |r_i^{(2)}|^{q-1}\)
for \(\beta = (\text{sign } \alpha) |\alpha|^{q-1}.\) We can determine the scalar \(\beta\) from the identity
\[<r^{(p)}, r^{(2)}> = <r^{(2)}, r^{(2)}>\] (5.6)
which holds for all \(p\) by Lemma 2.2(b). Using (5.3) and (5.6) we obtain
\[\beta = \frac{<r^{(2)}, r^{(2)}>}{<|r^{(2)}|^{q-1}, |r^{(2)}|^{p}>} = \frac{\sum_{i=1}^{m} |r_i^{(2)}|^2}{\sum_{i=1}^{m} |r_i^{(2)}|^q}.\] \(\square\)

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

**Theorem 5.2** Let \(A \in \mathbb{R}^{m \times n}_{m-1}, \ b \in \mathbb{R}^m\). Then there is a residual \(r^{(\infty)}\) satisfying
\[r^{(\infty)} = \beta \text{sign } r^{(2)},\] (5.7)
where
\[\beta = \frac{\sum_{i=1}^{m} |r_i^{(2)}|^2}{\sum_{i=1}^{m} |r_i^{(2)}|^q}.\] (5.8)

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

**Theorem 5.3** Let \(A \in \mathbb{R}^{m \times n}_{m-1}, \ b \in \mathbb{R}^m\). Then there is a residual \(r^{(1)}\) satisfying
\[r^{(1)} = \frac{\sum_{i=1}^{m} |r_i^{(2)}|^2}{r^2_{\ell}} e_\ell,\] (5.9)
where \(\ell\) is selected in \(\arg \max_i |r_i^{(2)}|\), and \(e_\ell\) denotes the \(\ell\)-th unit vector.
Proof. This follows from Lemma 2.2(b) and [8, Problem 3, p. 42]. □

From Theorems 5.1–5.3 it follows that the $\ell_p$–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 $A \in \mathbb{R}^{m \times n}$.

(a) Let $1 < p < \infty$. Then the $\ell_p$–approximate solutions of $Ax = b$ are exactly the solutions of the linear equation

$$Ax = b + \beta (\text{sign } r^{(2)}) \circ |r^{(2)}|^{q-1},$$

(5.10)

where $\beta$ is given by (5.4).

(b) Let $p = \infty$. Then any solution of (5.10), with $\beta$ given by (5.8) and $q = 1$, is an $\ell\infty$–approximate solution of $Ax = b$.

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

$$Ax = b + \sum_{i=1}^{m} \frac{|r_i^{(2)}|^2}{r_i^{(2)}} e_{\ell}$$

(5.11)

is an $\ell_1$–approximate solution of $Ax = b$. □

**Remark 5.1** For $A \in \mathbb{R}^{(n+1) \times n}$ 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**


