Hoffman's Error Bounds and Uniform Lipschitz Continuity of Best l(p) -Approximations

H. Berens

M. Finzel

W. Li
Old Dominion University

Y. Xu

Original Publication Citation

Berens, H., Finzel, M., Li, W., & Xu, Y. (1997). Hoffman's error bounds and uniform lipschitz continuity of best l_p-approximations. *Journal of Mathematical Analysis and Applications, 213*(1), 183-201. doi:10.1006/jmaa.1997.5521

JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS **213,** 183–201 (1997)
ARTICLE NO. AY975521

Hoffman's Error Bounds and Uniform Lipschitz Continuity of Best l_p-Approximations

H. Berens* and M. Finzel†

Mathematical Institute, University of Erlangen-Nuremberg, D-91054, Erlangen, Germany

W. Li‡

Department of Mathematics and Statistics, Old Dominion University, Norfolk, Virginia 23529-0077

and

Y. Xu§

Department of Mathematics, University of Oregon, Eugene, Oregon 97403-1222

Submitted by Joseph D. Ward

Received October 24, 1996

In a central paper on smoothness of best approximation in 1968 R. Holmes and B. Kripke proved among others that on $\mathbb{R}^n$, endowed with the l_p-norm, $1 < p < \infty$, the metric projection onto a given linear subspace is Lipschitz continuous where the Lipschitz constant depended on the parameter p. Using *Hoffman's Error Bounds* as a principal tool we prove uniform Lipschitz continuity of best l_p-approximations. As a consequence, we reprove and prove, respectively, Lipschitz continuity of the *strict best approximation* (**sba**, $p = \infty$) and of the *natural best approximation* (**nba**, $p = 1$). © 1997 Academic Press

* E-mail address: berens@mi.uni-erlangen.de.
† E-mail address: finzel@mi.uni-erlangen.de.
‡ E-mail address: wuli@math.odu.edu.
§ E-mail address: yuan@bright.uoregon.edu.

0022-247X/97 $25.00

1. INTRODUCTION

In the second week of July the four authors sat together in the Mathematical Institute in Erlangen to discuss various problems on approximation and got stuck on a problem of finite dimensional discrete linear approximation. For the second named author it is one of her research areas, and a recent exchange with the third named author fostered the discussion. But before we get into detail, we have to introduce the necessary notation.

Let $\mathbb{R}^n$ be the n-dimensional (Euclidean) space of column vectors. The p-norm of a vector x in $\mathbb{R}^n$ is denoted by $\|x\|_p := (\sum_{i=1}^n |x_i|^p)^{1/p}$ for $1 \le p < \infty$ and by $\|x\|_\infty := \max_{1 \le i \le n} |x_i|$ for $p = \infty$. For convenience, we use $\|\cdot\|$ as the 2-norm $\|\cdot\|_2$. For a subspace $\boldsymbol{U}$ of $\mathbb{R}^n$, the problem of best l_p-approximation of a vector x in $\mathbb{R}^n$ by elements of $\boldsymbol{U}$ is to find a vector u^* in $\boldsymbol{U}$ such that

$$\|x - u^*\|_p = \min\{\|x - u\|_p : u \in \boldsymbol{U}\}. \tag{1}$$

Let $\Pi_{\boldsymbol{U},p}(x)$ be the set of all best l_p-approximants of x in $\boldsymbol{U}$; *i.e.*,

$$\Pi_{\boldsymbol{U},p}(x) := \left\{u^* \in \boldsymbol{U} : \|x - u^*\|_p = \min_{u \in \boldsymbol{U}} \|x - u\|_p\right\}.$$

Then $\Pi_{\boldsymbol{U},p}$ is called the *metric projection* from $\mathbb{R}^n$ to $\boldsymbol{U}$ with respect to the p-norm.

For $1 < p < \infty$, $\Pi_{\boldsymbol{U},p}$ defines a continuous point-valued mapping because of the strict convexity of the norm, while for $p = 1$ or $p = \infty$, it is in general set-valued; *i.e.*, for $x \in \mathbb{R}^n$, $\Pi_{\boldsymbol{U},p}(x)$ might contain infinitely many elements. In fact, $\Pi_{\boldsymbol{U},p}(x)$ is a closed convex polytope. In this case, the metric projection $\Pi_{\boldsymbol{U},p}$ is Lipschitz continuous as a set-valued mapping with respect to the Hausdorff metric; *i.e.*, there exists a positive constant γ (depending only on $\boldsymbol{U}$) such that (cf. [17])

$$H\big(\Pi_{\boldsymbol{U},p}(x), \Pi_{\boldsymbol{U},p}(y)\big) \le \gamma \cdot \|x - y\| \qquad \text{for } x, y \in \mathbb{R}^n,\ p = 1 \text{ or } \infty; \tag{2}$$

here

$$H(S_1, S_2) := \max\left\{\max_{x \in S_1} \operatorname{dist}(x, S_2), \max_{y \in S_2} \operatorname{dist}(y, S_1)\right\}$$

denotes the *Hausdorff distance* of the sets S_1 and S_2 and $\operatorname{dist}(x, S_i) := \min\{\|x - v\| : v \in S_i\}$ the distance from the point x to the set S_i. Consequently, if $\Pi_{\boldsymbol{U},1}(x)$ (or $\Pi_{\boldsymbol{U},\infty}(x)$) contains exactly one element for every x in $\mathbb{R}^n$, then the metric projection $\Pi_{\boldsymbol{U},1}$ (or $\Pi_{\boldsymbol{U},\infty}$) is point-valued and Lipschitz continuous (as a mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$).

Since $\Pi_{U,p}(x)$ (with $p = 1$ or ∞) is a compact convex set for each x, it follows from (2) that one can use the Steiner point selector to get a Lipschitz continuous selection for $\Pi_{U,p}$ (cf. [2] and references therein). (Recall that a mapping $\sigma\colon \mathbb{R}^n \to U$ is called a selection for $\Pi_{U,p}$ if $\sigma(x) \in \Pi_{U,p}(x)$ for $x \in \mathbb{R}^n$.) However, the Steiner point selector requires information on the whole set $\Pi_{U,p}(x)$; thus it is not a practical way to construct a Lipschitz continuous selection for $\Pi_{U,p}$. The most intensively investigated selection for $\Pi_{U,\infty}$ is the *strict best approximation*, denoted by **sba**, that can be defined as the limit of $\Pi_{U,p}$ as $p \to \infty$:

$$\mathbf{sba}(x) := \lim_{p\to\infty} \Pi_{U,p}(x) \qquad \text{for } x \in \mathbb{R}^n. \tag{3}$$

The strict best approximation was introduced by Rice as the best of best approximations. The limit (3) was first proved by Descloux and was rediscovered later by Mityagin [19]. For the computation of the strict best approximation, see [9, 1, 3]. Even though **sba** was believed to be a continuous mapping, a formal proof based on generalized inverses first appeared in [10]; another proof can be found in [5]. The limit (3) gives an easy description of **sba**(x), but it does not reveal much about the properties and structure of strict best approximation. Recently, Finzel [6] used Plücker-Grassmann coordinates to give a complete structural description of **sba**. As a consequence, she proved that $\mathbb{R}^n$ can be subdivided into finitely many polyhedral cones where on each of them **sba** is a linear mapping and, hence, **sba** is Lipschitz continuous. For further properties and references on strict best approximation, see [6, 13].

The selection for $\Pi_{U,1}$ corresponding to **sba** is the *natural best approximation*, denoted by **nba**, which is defined by the limit of $\Pi_{U,p}$ as $p \to 1^+$ [15]:

$$\mathbf{nba}(x) := \lim_{p\to 1^+} \Pi_{U,p}(x) \qquad \text{for } x \in \mathbb{R}^n. \tag{4}$$

Landers and Rogge [15] proved that for each x, $\lim_{p\to 1^+} \Pi_{U,p}(x)$ exists (*i.e.*, **nba**(x) is well-defined). Independently, Fischer [8] also proved the existence of this limit. In addition, Landers and Rogge [15] characterized **nba**(x) as the unique solution of the following minimization problem:

$$\min\left\{ \sum_{i=1}^{n} |x_i - u_i| \cdot \log|x_i - u_i| \colon u \in \Pi_{U,1}(x) \right\} \qquad \text{for } x \in \mathbb{R}^n. \tag{5}$$

We should point out that in 1921, Jackson already established the limit relation (4) for the median (corresponding to the case dim $U = 1$). See [4, 7] for more references on the natural best approximation. However, the objective function in (5) is not a differentiable function and, hence, there is no standard approach for studying stability properties, in our case, Lipschitz continuity.

Our approach to establish the Lipschitz continuity of **sba** and **nba**, without studying the structure of the mappings directly, is to prove the uniform Lipschitz continuity of $\Pi_{U,p}$ with respect to $1 < p < \infty$, which is the main aim of the paper. The result is summarized in the following theorem.

THEOREM 1. *Let U be a linear subspace of $\mathbb{R}^n$. Then there exists a positive constant λ such that*

$$\left\|\Pi_{U,p}(x) - \Pi_{U,p}(y)\right\| \leq \lambda \cdot \|x-y\|, \qquad \text{for } x, y \in \mathbb{R}^n, 1 < p < \infty. \quad (6)$$

As a consequence, the strict best approximation **sba** *and the natural best approximation* **nba** *are Lipschitz continuous selections for $\Pi_{U,\infty}$ and $\Pi_{U,1}$, respectively.*

The theorem was conjectured in [15] in connection with a general discussion on best approximation in polyhedral spaces, which was motivated by the paper of Holmes and Kripke [12] on smoothness of best approximation in Banach spaces. In [12], Holmes and Kripke proved that $\Pi_{U,p}$ is Lipschitz continuous for each $p \in (1, \infty)$. Their proof was given for $L_p(T, \mu)$ with a nonatomic measure μ on a compact Hausdorff space T, and because of that, they had to restrict their arguments to $2 < p < \infty$. But for $\mathbb{R}^n$ endowed with p-norms, their proof holds true for $1 < p < 2$ as well. That is, for $1 < p < \infty$,

$$\left\|\Pi_{U,p}(x) - \Pi_{U,p}(y)\right\| \leq \lambda_p \cdot \|x-y\|, \qquad \text{for } x, y \in \mathbb{R}^n, \quad (7)$$

where λ_p is some positive constant depending on U and p only.

Technically, (6) is an improvement of (7). The improvement comes from the analysis of the Jacobian of $\Pi_{U,p}$, which led us to the following theorem about matrix inequalities.

THEOREM 2. *Let B be an $n \times n$ matrix. Then there exists a positive constant λ such that*

$$\|WBy\| \leq \lambda \cdot \|BB^+WBy\| \qquad \text{for } y \in \mathbb{R}^m \text{ and } W = \operatorname{diag}(w_1, \ldots, w_n) \text{ with } w_i \geq 0, \quad (8)$$

where B^+ is the pseudo-inverse of B and $\operatorname{diag}(w_1, \ldots, w_n)$ the diagonal matrix whose ith diagonal entry is w_i.

If we define U to be the column space of B; *i.e.*, $U := \{By\colon\ y \in \mathbb{R}^m\}$, then (8) can be interpreted as follows: the norm of a scaled vector in U is uniformly bounded by the norm of the orthogonal projection of the scaled vector to U. Hoffman's error bound, see [11], for approximate solutions of linear inequalities and equalities turns out to be the key in establishing (8). The reason is that the set generated by scaling vectors in U is a union of finitely many polyhedral sets. In addition, we can even give an explicit estimate of λ in (6) in terms of submatrices of any matrix whose columns form an orthonormal basis of U. For further applications of Hoffman's error bound in connection with best approximation in polyhedral spaces see [17].

The paper is organized as follows. Section 2 is devoted to the study of the matrix inequality (8) and its ramifications. In Section 3, we establish an estimate of the Jacobian of $\Pi_{U,p}$ based on (8). Then we give two proofs of Theorem 1: one is self-contained and another is a simpler proof based on Lipschitz continuity of $\Pi_{U,p}$ proved by Holmes and Kripke.

To conclude this section we introduce commonly used notations in this paper. For a matrix A (or a vector x) and an index set J, we define A_J (or x_J) to be the matrix (or the vector) consisting of rows of A (or components of x) whose corresponding indices are in J. In particular, A_i (or x_i) represents the ith row (or ith component) of A (or x). For an index set $J \subset \{1,\dots,n\}$, let $J^c := \{1,\dots,n\} \setminus J$ be the complement of J. The dot product in $\mathbb{R}^n$ is denoted by $\langle x, y\rangle = \sum_{i=1}^n x_i y_i$ for $x, y \in \mathbb{R}^n$. For an $n \times m$ matrix B, its spectral norm is defined by $\|B\| := \max\{\|By\|\colon y \in \mathbb{R}^m$ with $\|y\| \le 1\}$. The $k \times k$ identity matrix is denoted by I_k and sometimes we also use I as the identity matrix if there is no confusion about its dimension. For the relative interior of a subset K of $\mathbb{R}^n$, we will use the symbol $\mathrm{ri}(K)$. Finally, a subset K of $\mathbb{R}^n$ is called a polyhedral set if it is an intersection of finitely many closed half-spaces.

2. HOFFMAN'S ERROR BOUND AND MATRIX INEQUALITY

As preparation of the proof of Theorem 2 we formulate and prove four lemmas.

LEMMA 3. *Let B be an $n \times m$ matrix. Then*

$$\langle x - BB^+x, By\rangle = 0 \qquad \text{for } x \in \mathbb{R}^n,\ y \in \mathbb{R}^m. \tag{9}$$

Proof. It is well known that B^+x is the least norm solution of the following least squares problem:

$$\min_{y\in\mathbb{R}^m} \tfrac{1}{2}\|x - By\|^2. \tag{10}$$

By a characterization for an optimal solution of (10), we have $B^T(x - BB^+x) = 0$, which implies

$$\langle x - BB^+x, By\rangle = \left\langle B^T(x - BB^+x), y\right\rangle = 0 \qquad \text{for } x \in \mathbb{R}^n,\ y \in \mathbb{R}^m. \quad \blacksquare$$

LEMMA 4. *Suppose that B is an $n \times m$ matrix, D a symmetric positive semidefinite matrix, and $y \in \mathbb{R}^m$. If $BB^+DBy = 0$, then $DBy = 0$.*

Proof. Let $u = By$ and $x = Du$. Then $BB^+x = 0$. By (9) in Lemma 3, we have

$$u^T Du = \langle Du, u \rangle = \langle x, By \rangle = \langle x - BB^+x, By \rangle = 0.$$

Since D is symmetric and positive semidefinite, there exists a matrix Q such that $D = Q^T Q$. Thus, $\|Qu\|^2 = u^T Du = 0$; *i.e.*, $Qu = 0$. So $DBy = Du = Q^T(Qu) = 0$. ∎

LEMMA 5. *If X is a subset of $\mathbb{R}^n$, then the set*

$$\mathsf{V} := \{Wx\colon x \in X \text{ and } W = \operatorname{diag}(w_1, \ldots, w_n) \text{ with } w_i \geq 0\} \tag{11}$$

is a union of finitely many convex polyhedral sets.

Proof. For any $\epsilon = (\epsilon_1, \ldots, \epsilon_n)$ with $\epsilon_i \in \{-1, 0, 1\}$, define the octant corresponding to ϵ, denoted by $\operatorname{Oct}(\epsilon)$, as

$$\operatorname{Oct}(\epsilon) := \{x \in \mathbb{R}^n\colon \epsilon_i x_i \geq 0 \quad \text{for } \epsilon_i \neq 0, \text{ and } x_i = 0 \text{ for } \epsilon_i = 0\}.$$

Let

$$\mathsf{I}_X := \{\epsilon = (\epsilon_1, \ldots, \epsilon_n) \in \{-1, 0, 1\}^n\colon X \cap \operatorname{ri}[\operatorname{Oct}(\epsilon)] \neq \emptyset\}.$$

First we claim that

$$\mathsf{V} \subset \bigcup_{\epsilon \in \mathsf{I}_X} \operatorname{Oct}(\epsilon). \tag{12}$$

Let $x \in X$ and define $\epsilon = \operatorname{sign}(x)$, where $\operatorname{sign}(x) = (\operatorname{sign}(x_1), \ldots, \operatorname{sign}(x_n))$. Then

$$\operatorname{Oct}(\epsilon) = \{Wx\colon W = \operatorname{diag}(w_1, \ldots, w_n) \text{ with } w_i \geq 0\}.$$

Moreover, $x \in X \cap \operatorname{ri}[\operatorname{Oct}(\epsilon)]$. So $\epsilon \in \mathsf{I}_X$, which implies

$$\mathsf{V} = \bigcup_{x \in X} \{Wx\colon W = \operatorname{diag}(w_1, \ldots, w_n) \text{ with } w_i \geq 0\} \subset \bigcup_{\epsilon \in \mathsf{I}_X} \operatorname{Oct}(\epsilon),$$

proving (12).

Next we prove that $\operatorname{Oct}(\epsilon) \subset \mathsf{V}$ for $\epsilon \in \mathsf{I}_X$. Let $\epsilon \in \mathsf{I}_X$. Then there exists $x \in X$ such that $\operatorname{sign}(x) = \epsilon$. (Note that $x \in \operatorname{ri}[\operatorname{Oct}(\epsilon)]$ if and only if

$\operatorname{sign}(x) = \epsilon$.) Let $v \in \operatorname{Oct}(\epsilon)$. Define

$$w_i = \begin{cases} v_i/x_i & \text{if } x_i \neq 0 \\ 0 & \text{if } x_i = 0. \end{cases}$$

Since $\operatorname{sign}(v_i) = \epsilon_i = \operatorname{sign}(x_i)$ when $v_i \neq 0$, we have $w_i \geq 0$ for $1 \leq i \leq n$. Note that $v_i = w_i x_i$ for $1 \leq i \leq n$; *i.e.*, $v = Wx$ with $W = \operatorname{diag}(w_1, \dots, w_n)$. So $v \in \mathcal{V}$ and therefore,

$$\mathcal{V} = \bigcup_{x \in X} \{Wx \colon W = \operatorname{diag}(w_1, \dots, w_n) \text{ with } w_i \geq 0\} \supset \bigcup_{\epsilon \in \mathcal{I}_X} \operatorname{Oct}(\epsilon).$$

Since $\operatorname{Oct}(\epsilon)$ is a convex polyhedral set and $\mathcal{I}_X$ finite, $\mathcal{V}$ is a union of finitely many convex polyhedral sets. ∎

LEMMA 6 (A Variation of Hoffman's Error Bound). *Let X be a polyhedral subset of $\mathbb{R}^n$ and $S := \{x \in X \colon Ax = b\}$. Then there exists a positive constant λ (depending only on X and A) such that*

$$\operatorname{dist}(x, S) \leq \lambda \cdot \|Ax - b\| \qquad \text{for } x \in X. \tag{13}$$

Proof. Since X is a convex polyhedral set, there exist a matrix C and a vector d such that $X = \{x \in \mathbb{R}^n \colon Cx \leq d\}$. Then by Hoffman's error bound for approximate solutions of systems of linear inequalities [11], there exists a positive constant λ such that

$$\operatorname{dist}(x, S) \leq \lambda \cdot \left(\|Ax - b\| + \|(Cx - d)_+\|\right) \qquad \text{for } x \in \mathbb{R}^n, \tag{14}$$

where z_+ denotes the vector whose ith component is $\max\{z_i, 0\}$. Note that $x \in X$ if and only if $(Cx - d)_+ = 0$. Thus, (13) follows from (14). ∎

Proof of Theorem 2. We claim that for an $n \times m$ matrix B there exists a positive constant λ such that

$$\|WBy\| \leq \lambda \cdot \|BB^+ WBy\| \qquad \text{for } y \in \mathbb{R}^m \text{ and } W = \operatorname{diag}(w_1, \dots, w_n) \text{ with } w_i \geq 0. \tag{15}$$

Let $X := \{By \colon y \in \mathbb{R}^m\}$ and

$$\mathcal{V} := \left\{Wx \colon x \in X \text{ and } W = \operatorname{diag}(w_1, \dots, w_n) \text{ with } w_i \geq 0\right\}.$$

Then by Lemma 5 the set $\mathcal{V}$ is a union of finitely many convex polyhedral sets $\{V_1, V_2, \dots, V_k\}$; *i.e.*,

$$\mathcal{V} = \bigcup_{i=1}^{k} V_i.$$

By Lemma 6, there exist positive constants λ_i such that

$$\operatorname{dist}(v, S_i) \le \lambda_i \cdot \|BB^+ v\| \qquad \text{for } v \in V_i, \tag{16}$$

where

$$S_i := \{z \in V_i\colon BB^+ z = 0\}.$$

Let $z \in S_i$. Since $z \in V_i \subset \mathcal{V}$, there exist $y \in \mathbb{R}^m$ and a diagonal matrix $W = \operatorname{diag}(w_1, \ldots, w_n)$ with $w_i \ge 0$ such that $z = WBy$. Since $z \in S_i$, we have $BB^+ WBy = BB^+ z = 0$. Since W is symmetric and positive semidefinite, it follows from Lemma 4 that $z = WBy = 0$. Therefore, $S_i = \{0\}$ and $\operatorname{dist}(v, S_i) = \|v\|$. The inequality (16) actually reads

$$\|v\| \le \lambda_i \cdot \|BB^+ v\| \qquad \text{for } v \in V_i. \tag{17}$$

Setting $\lambda := \max\{\lambda_i\colon 1 \le i \le k\}$, we obtain

$$\|v\| \le \lambda \cdot \|BB^+ v\| \qquad \text{for } v \in \mathcal{V}. \tag{18}$$

By the definition of $\mathcal{V}$, we know that the inequality in Theorem 2 holds. ∎

Remark. If we allow a bigger motion of the subset X than multiplication with a positive diagonal matrix, then Theorem 2 does not hold any more. To be precise, let D be a symmetric positive semidefinite matrix, $X := \{DBy\colon y \in \mathbb{R}^m\}$, and $S := \{x \in X\colon BB^+ x = 0\}$. Then, by Lemma 5, $S = \{0\}$. It follows from Lemma 6 that there exists a positive constant $\lambda(D)$ such that

$$\operatorname{dist}(x, S) \le \lambda(D) \cdot \|BB^+ x\| \qquad \text{for } x \in X;$$

i.e.,

$$\|DBy\| \le \lambda(D) \cdot \|BB^+ DBy\| \qquad \text{for } y \in \mathbb{R}^m. \tag{19}$$

Thus, in the light of Theorem 2, one might expect that the $\lambda(D)$ in (19) are uniformly bounded for all symmetric positive semidefinite matrices. However, the following example shows that the $\lambda(D)$ in (19) are not uniformly bounded with respect to all symmetric positive semidefinite matrices.

Let $n = 2$, $m = 1$, $y = 1 \in \mathbb{R}$, $B = \binom{1}{1}$, $\pi/4 < \alpha < \pi/2$,

$$Q = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}, \qquad W = \begin{pmatrix} w & 0 \\ 0 & 0 \end{pmatrix}, \qquad w = \frac{\sqrt{2}}{\sin\alpha - \cos\alpha},$$

and $D = Q^T W Q$. Then Q is the orthogonal matrix that represents rotation by the angle α and $D = Q^T W Q$ is a symmetric positive semidefinite matrix. By matrix multiplications, we obtain

$$DBy = Q^T WQBy = \sqrt{2} \begin{pmatrix} -\cos\alpha \\ \sin\alpha \end{pmatrix}.$$

Since $B^+ = (B^TB)^{-1}B^T = \frac{1}{2}(1\,1)$, we can compute

$$BB^+DBy = \frac{1}{\sqrt{2}}\begin{pmatrix} \sin\alpha - \cos\alpha \\ \sin\alpha - \cos\alpha \end{pmatrix}.$$

So

$$\lambda(D) \geq \frac{\|DBy\|}{\|BB^+DBy\|} = \frac{\sqrt{2}}{\sin\alpha - \cos\alpha}. \tag{20}$$

Therefore, $\lambda(D)$ is unbounded as $\alpha \to (\pi/4)^+$. As a consequence, Theorem 2 fails to be true if we allow W to be an arbitrary symmetric positive semidefinite matrix.

COROLLARY 7. *Let Q be an $n \times m$ matrix with rank m. Then there exists a positive constant γ such that*

$$\|WQy\| \leq \gamma \cdot \|Q^TWQy\| \qquad \textit{for } y \in \mathbb{R}^m \textit{ and } W = \operatorname{diag}(w_1, \dots, w_n) \textit{ with } w_i \geq 0. \tag{21}$$

Proof. Since the columns of Q are linearly independent, $Q^+ = (Q^TQ)^{-1}Q^T$. Thus, by Theorem 2, there exists a positive constant λ such that

$$\|WQy\| \leq \lambda \cdot \left\| Q\left(Q^TQ\right)^{-1}Q^TWQy \right\|$$
$$\text{for } y \in \mathbb{R}^m \text{ and } W = \operatorname{diag}(w_1, \dots, w_n) \text{ with } w_i \geq 0.$$

Since $\|Q(Q^TQ)^{-1}z\| \leq \|Q(Q^TQ)^{-1}\| \cdot \|z\|$, the above inequality implies

$$\|WQy\| \leq \gamma \cdot \|Q^TWQy\|$$
$$\text{for } y \in \mathbb{R}^m \text{ and } W = \operatorname{diag}(w_1, \dots, w_n) \text{ with } w_i \geq 0,$$

where $\gamma = \lambda\|Q(Q^TQ)^{-1}\|$. ∎

Remark. By using explicit estimates of Lipschitz constants for feasible solutions of a system of linear equalities and inequalities, we can get explicit estimates of λ in Theorem 2 and γ in Corollary 7. For illustration purpose, we derive the following estimate for γ,

$$\|WQy\| \leq \bar{\gamma} \cdot \|Q^TWQy\|$$
$$\text{for } y \in \mathbb{R}^m \text{ and } W = \operatorname{diag}(w_1, \dots, w_n) \text{ with } w_i \geq 0, \tag{22}$$

where

$$\overline{\gamma} := \max_J \left\{ \left\| \begin{pmatrix} I_{n-m} & -Q_{J^c} Q_J^{-1} \\ 0 & Q_J^{-1} \end{pmatrix} \right\| : Q_J \text{ is a nonsingular } m \times m \text{ matrix} \right\}.$$

Let $X := \{Qy \colon y \in \mathbb{R}^m\}$ and let V, $\mathrm{Oct}(\epsilon)$, I_X be defined as in the proof of Lemma 5. Let $\epsilon = (\epsilon_1, \ldots, \epsilon_n) \in I_X$ and consider the following system of equalities and inequalities,

$$Q^T x = b \qquad \text{and} \qquad Ax \leq 0, \tag{23}$$

where $b \in \mathbb{R}^m$ and

$$A = \begin{pmatrix} -I_{J_+} \\ I_{J_-} \\ I_{J_0} \\ -I_{J_0} \end{pmatrix},$$

I is the $n \times n$ identity matrix, $J_+ := \{i \colon \epsilon_i = 1\}$, $J_- := \{i \colon \epsilon_i = -1\}$, and $J_0 := \{i \colon \epsilon_i = 0\}$. (Note that $Ax \leq 0$ if and only if $x \in \mathrm{Oct}(\epsilon)$.) Let $S(b)$ be the solution set of (23). By the remark on page 34 in [18], we have

$$H(S(b), S(b')) \leq \tilde{\gamma} \cdot \|b - b'\| \qquad \text{for } b, b' \in \mathbb{R}^m, \tag{24}$$

where

$$\tilde{\gamma} := \max_J \left\{ \left\| \begin{pmatrix} A_{\bar{J}} \\ Q^T \end{pmatrix}^{-1} \right\| : \begin{pmatrix} A_{\bar{J}} \\ Q^T \end{pmatrix} \text{ is a nonsingular } n \times n \text{ matrix} \right\}.$$

For any $x \in \mathrm{Oct}(\epsilon)$, let $b = Q^T x$ and $b' = 0$. Then $S(b') = \{0\}$ (by Lemma 4 and $Q^+ = (Q^T Q)^{-1} Q^T$) and $x \in S(b)$. So (24) implies

$$\|x\| \leq H(S(b), S(b')) \leq \tilde{\gamma} \cdot \|b - b'\| = \tilde{\gamma} \cdot \|Q^T x\| \qquad \text{for } x \in \mathrm{Oct}(\epsilon),\ \epsilon \in I_X. \tag{25}$$

Since the rows of $A_{\bar{J}}$ are rows of I or $-I$, let $D = \mathrm{diag}(d_1, \ldots, d_n)$ be such that $d_i = \pm 1$ and

$$D \begin{pmatrix} A_{\bar{J}} \\ Q^T \end{pmatrix} = \begin{pmatrix} I_{J^c} \\ Q^T \end{pmatrix},$$

where J is some index subset of $\{1,\dots,n\}$. Since D is an orthogonal matrix, we have

$$\begin{aligned}\left\|\begin{pmatrix}A_{\bar J}\\ Q^T\end{pmatrix}^{-1}\right\| &= \left\|\left(D\begin{pmatrix}A_{\bar J}\\ Q^T\end{pmatrix}\right)^{-1}\right\|\\ &= \left\|\begin{pmatrix}I_{J^c}\\ Q^T\end{pmatrix}^{-1}\right\| = \left\|\begin{pmatrix}I_{n-m} & 0\\ Q_{J^c}^T & Q_J^T\end{pmatrix}^{-1}\right\|\\ &= \left\|\begin{pmatrix}I_{n-m} & Q_{J^c}\\ 0 & Q_J\end{pmatrix}^{-1}\right\| = \left\|\begin{pmatrix}I_{n-m} & -Q_{J^c}Q_J^{-1}\\ 0 & Q_J^{-1}\end{pmatrix}\right\|,\end{aligned}$$

where the third equality holds since we only exchange columns of the matrix

$$\begin{pmatrix}I_{J^c}\\ Q^T\end{pmatrix}^{-1},$$

the fourth one follows from $\|B\| = \|B^T\|$, and the last equality is derived from the calculation of the inverse of $\begin{pmatrix}I_{n-m} & Q_{J^c}\\ 0 & Q_J\end{pmatrix}$. Thus, (25) is equivalent to (22).

3. UNIFORM LIPSCHITZ CONTINUITY OF BEST l_p-APPROXIMATION

For $x \in \mathbb{R}^n$, consider the best l_p-approximation problem of finding $\Pi_{U,p}(x)$ in a subspace U of $\mathbb{R}^n$ such that

$$\left\|x - \Pi_{U,p}(x)\right\|_p = \min\left\{\|x-u\|_p : u \in U\right\}, \qquad 1 < p < \infty.$$

As we pointed out in the Introduction, Holmes and Kripke actually proved that for each $p \in (1,\infty)$, there exists a positive constant λ_p such that

$$\left\|\Pi_{U,p}(x) - \Pi_{U,p}(y)\right\|_p \le \lambda_p \cdot \|x-y\|_p \qquad \text{for } x, y \in \mathbb{R}^n. \tag{26}$$

We want to show that λ_p in (26) is uniformly bounded for $1 < p < \infty$. We start with an estimate on the Jacobian of the metric projection $\Pi_{U,p}$.

THEOREM 8. *Let U be a subspace of $\mathbb{R}^n$. Then there exists a positive constant γ such that*

$$\left\|\nabla_x \Pi_{U,p}(x)\right\| \le \gamma \qquad \textit{for } 1 < p < \infty,\ x \in D_p, \tag{27}$$

where

$$\mathsf{D}_p := \left\{ x \in \mathbb{R}^n \colon \left(x - \Pi_{\mathsf{U},p}(x) \right)_i \neq 0 \qquad \textit{for } 1 \leq i \leq n \right\}. \tag{28}$$

Proof. Let U be a matrix whose columns $\{u^1, \ldots, u^m\}$ form an orthonormal basis of a subspace U of $\mathbb{R}^n$. Since $\Pi_{\mathsf{U},p}(x)$ is unique for each $1 < p < \infty$, there exists a unique vector

$$t_p(x) = \begin{pmatrix} t_{1,p}(x) \\ \vdots \\ t_{m,p}(x) \end{pmatrix}$$

such that

$$\Pi_{\mathsf{U},p}(x) = \sum_{k=1}^{m} t_{k,p}(x) u^k = U t_p(x). \tag{29}$$

Let the function $f_p \colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be defined by

$$f_p(x,t) := \frac{1}{p} \left\| x - \sum_{k=1}^{m} t_k u^k \right\|_p^p. \tag{30}$$

Then for each fixed x, the gradient of f_p with respect to t has to be zero at $t_p(x)$; *i.e.*, $\nabla_t f_p(x, t_p(x)) = 0$. Thus, for any given x, $t_p(x)$ is the unique solution of the following system of nonlinear equations:

$$\sum_{r=1}^{n} \left| x_r - \sum_{k=1}^{m} t_k u_r^k \right|^{p-1} \operatorname{sign}\left(x_r - \sum_{k=1}^{m} t_k u_r^k \right) u_r^i = 0, \qquad \text{for } 1 \leq i \leq m. \tag{31}$$

Let $\bar{x} \in \mathsf{D}_p$. Then $\nabla_t f_p(x,t)$ is continuously differentiable with respect to (x,t) in a neighborhood of $(\bar{x}, t_p(\bar{x}))$. In fact,

$$\frac{\partial^2 f_p(x,t)}{\partial t_i\, \partial t_j} = (p-1) \sum_{r=1}^{n} \left| x_r - \sum_{k=1}^{m} t_k u_r^k \right|^{p-2} u_r^i u_r^j, \qquad \text{for } 1 \leq i, j \leq m, \tag{32}$$

and

$$\frac{\partial^2 f_p(x,t)}{\partial t_i\, \partial x_j} = (1-p) \left| x_j - \sum_{k=1}^{m} t_k u_j^k \right|^{p-2} u_j^i, \qquad \text{for } 1 \leq i \leq m, 1 \leq j \leq n. \tag{33}$$

Equivalently, we have the following matrix representations of the Jacobians $\nabla^2_{tt} f_p(x,t)$ and $\nabla^2_{tx} f_p(x,t)$, respectively,

$$\nabla^2_{tt} f_p(x,t) = (p-1)U^T W U \qquad \text{in } \mathbb{R}^{m\times m}, \tag{34}$$

$$\nabla^2_{tx} f_p(x,t) = (1-p)U^T W \qquad \text{in } \mathbb{R}^{m\times n}, \tag{35}$$

where $W = \operatorname{diag}(w_1,\dots,w_n)$ and $w_i := |x_i - \sum_{k=1}^m t_k u_i^k|^{p-2}$. Since $w_i > 0$ for (x,t) in a neighborhood $\mathcal{O}(\bar{x}, t_p(\bar{x}))$ of $(\bar{x}, t_p(\bar{x}))$, the Jacobian $\nabla^2_{tt} f_p(x,t)$ is a positive definite matrix and continuous in $\mathcal{O}(\bar{x}, t_p(\bar{x}))$ with respect to (x,t). By the implicit function theorem, t_p is a differentiable function of x in a neighborhood $\mathcal{O}(\bar{x})$ of $\bar{x}$. Moreover,

$$\begin{aligned}\nabla_x t_p(x) &= -\left[\nabla^2_{tt} f_p\big(x, t_p(x)\big)\right]^{-1} \nabla^2_{tx} f_p\big(x, t_p(x)\big)\\ &= -\left[(p-1)U^T W U\right]^{-1}(1-p)U^T W = \left[U^T W U\right]^{-1} U^T W. \end{aligned}\tag{36}$$

The gradient is an $m \times n$ matrix whose (i,j)th entry is $\partial t_{i,p}(x)/\partial x_j$. Thus, we obtain

$$\begin{aligned}\left\| \nabla_x t_p(x) \right\| &= \left\| \left[\nabla_x t_p(x)\right]^T \right\| = \sup_{z\neq 0} \frac{\left\| \left[\nabla_x t_p(x)\right]^T z \right\|}{\|z\|}\\ &= \sup_{y\neq 0} \frac{\left\| \left[\nabla_x t_p(x)\right]^T U^T W U y \right\|}{\|U^T W U y\|} = \sup_{y \neq 0} \frac{\|WUy\|}{\|U^T W U y\|} \le \gamma, \end{aligned}\tag{37}$$

where the first equality follows from the definition of the norm of an adjoint matrix, the second one is by the definition of the spectral norm, the third one follows from the nonsingularity of $U^T W U$, the fourth is derived from (36), and finally, the last inequality follows from Corollary 7. Since $\Pi_{\mathcal{U},p}(x) = U t_p(x)$, by the chain rule, we get $\nabla_x \Pi_{\mathcal{U},p}(x) = U \nabla_x t_p(x)$ for $x \in \mathcal{O}(\bar{x})$. Since $\|U\| = 1$,

$$\left\| \nabla_x \Pi_{\mathcal{U},p}(\bar{x}) \right\| = \left\| U \nabla_x t_p(\bar{x}) \right\| \le \|U\| \cdot \left\| \nabla_x t_p(\bar{x}) \right\| = \left\| \nabla_x t_p(\bar{x}) \right\| \le \gamma,$$

where γ is independent of $\bar{x} \in \mathcal{D}_p$ and $p \in (1,\infty)$. ∎

Remark. It follows from the remark after Corollary 7 and (37) that if U is an $n \times m$ matrix whose columns form an orthonormal basis of $\mathcal{U}$, then

$$\left\| \nabla_x \Pi_{\mathcal{U},p}(\bar{x}) \right\| \le \bar{\gamma} \qquad \text{for } x \in \mathcal{D}_p, \tag{38}$$

where

$$\bar{\gamma} := \max_J \left\{ \left\| \begin{pmatrix} I_{n-m} & -U_{J^c} U_J^{-1} \\ 0 & U_J^{-1} \end{pmatrix} \right\| : U_J \text{ is a nonsingular } m \times m \text{ matrix} \right\}.$$

Proof of Theorem 1. Now we are ready to prove Theorem 1: If U is a subspace of $\mathbb{R}^n$, then there exists a positive constant λ such that

$$\left\|\Pi_{U,p}(x) - \Pi_{U,p}(y)\right\| \leq \lambda \cdot \|x-y\| \qquad \text{for } x, y \in \mathbb{R}^n, 1 < p < \infty. \tag{39}$$

We prove the theorem by induction. Obviously, if $\dim U = 0$, then (39) holds. We make the inductive hypothesis that (39) holds for any n and any subspace U of dimension less than m. Now assume that U is an m-dimensional subspace of $\mathbb{R}^n$.

Let $J := \{j\colon \text{there exists } u \in U \text{ such that } u_j \neq 0\}$ and let $u \in U$. Then $u_i = 0$ for $i \notin J$. If $U_J := \{u_J\colon u \in U\}$ denotes the restriction of U to the J components, $[\Pi_{U,p}(x)]_J = \Pi_{U_J,p}(x_J)$. Therefore, if there exists a positive constant λ such that

$$\left\|\Pi_{U_J,p}(x_J) - \Pi_{U_J,p}(y_J)\right\| \leq \lambda \cdot \|x_J - y_J\| \qquad \text{for } x, y \in \mathbb{R}^n, 1 < p < \infty,$$

then (39) holds. Thus we may replace U by U_J and assume that there exists $u \in U$ such that $u_i \neq 0$ for $1 \leq i \leq n$.

Let $U^{(i)} := \{u \in U\colon u_i = 0\}$, $1 \leq i \leq n$. Then $\dim U^{(i)} < \dim U$ (since $U \setminus U^{(i)} \neq \emptyset$). By our inductive hypothesis, there exists a positive constant β such that

$$\left\|\Pi_{U^{(i)},p}(x) - \Pi_{U^{(i)},p}(y)\right\| \leq \beta \cdot \|x-y\| \qquad \text{for } x, y \in \mathbb{R}^n, 1 < p < \infty. \tag{40}$$

Define

$$D_p := \left\{x \in \mathbb{R}^n\colon \left[x - \Pi_{U,p}(x)\right]_i \neq 0 \qquad \text{for } 1 \leq i \leq n\right\},$$

and for $1 \leq i \leq n$

$$D_{p,i} := \left\{x \in \mathbb{R}^n\colon \left[x - \Pi_{U,p}(x)\right]_i = 0\right\}.$$

By Theorem 8, there exists a positive constant γ such that (27) holds. Then it follows that

$$\begin{aligned}\left\|\Pi_{U,p}(x) - \Pi_{U,p}(y)\right\| &\leq \gamma \cdot \|x-y\|, \\ &\text{whenever } 1 < p < \infty \text{ and } x, y \in D_p \text{ with } \theta x + (1-\theta) y \in D_p \\ &\text{for } 0 \leq \theta \leq 1.\end{aligned} \tag{41}$$

This proves the desired estimate locally on D_p with the global Lipschitz constant γ. To prove (39), *i.e.*, to prove Lipschitz continuity of the metric projection on $\mathbb{R}^n$ uniformly with respect to $1 < p < \infty$, we first have to investigate the behavior of $\Pi_{\mathcal{U},p}(x)$ for $x \in \mathcal{D}_{p,i}$, $1 \le i \le n$. Let U be an $n \times m$ matrix whose columns form a basis of $\mathcal{U}$ and let $t_p(x) \in \mathbb{R}^m$ be such that

$$\Pi_{\mathcal{U},p}(x) = Ut_p(x). \tag{42}$$

Then, for $x \in \mathcal{D}_{p,i}$, we have

$$U_i t_p(x) = \left[\Pi_{\mathcal{U},p}(x)\right]_i = x_i. \tag{43}$$

Since there exists $u \in \mathcal{U}$ such that $u_i \neq 0$, U_i has to be a nonzero row. Let

$$\bar{t}_p = U_i^T\left(U_i U_i^T\right)^{-1} x_i \qquad \text{and} \qquad \tilde{t}_p = t_p(x) - \bar{t}_p. \tag{44}$$

Then

$$U_i \tilde{t}_p = U_i t_p(x) - U_i \bar{t}_p = x_i - x_i = 0. \tag{45}$$

That is, $U\tilde{t}_p \in \mathcal{U}^{(i)}$. Moreover, for any $u \in \mathcal{U}^{(i)}$,

$$\begin{aligned}\left\|\left(x - U\bar{t}_p\right) - u\right\|_p &= \left\|x - \left(U\bar{t}_p + u\right)\right\|_p \ge \left\|x - \Pi_{\mathcal{U},p}(x)\right\|_p \\ &= \left\|\left(x - U\bar{t}_p\right) - U\tilde{t}_p\right\|_p.\end{aligned}$$

Thus,

$$\Pi_{\mathcal{U}^{(i)},p}\left(x - U\bar{t}_p\right) = U\tilde{t}_p. \tag{46}$$

It follows from (42), the second equality in (44), and (46) that

$$\Pi_{\mathcal{U},p}(x) = U\bar{t}_p + \Pi_{\mathcal{U}^{(i)},p}\left(x - U\bar{t}_p\right). \tag{47}$$

From the above identity and the first equality in (44) we derive the following explicit representation of $\Pi_{\mathcal{U},p}(x)$ in terms of $\Pi_{\mathcal{U}^{(i)},p}$:

$$\begin{aligned}\Pi_{\mathcal{U},p}(x) &= UU_i^T\left(U_i U_i^T\right)^{-1} x_i + \Pi_{\mathcal{U}^{(i)},p}\left(x - UU_i^T\left(U_i U_i^T\right)^{-1} x_i\right) \\ &\quad \text{for } x \in \mathcal{D}_{p,i}.\end{aligned} \tag{48}$$

It follows from (40) and (48) that there is a positive constant η, independent of $1 < p < \infty$, such that

$$\left\|\Pi_{\mathcal{U},p}(x) - \Pi_{\mathcal{U},p}(y)\right\| \le \eta \cdot \|x - y\| \qquad \text{for } x, y \in \mathcal{D}_{p,i},\ 1 \le i \le n. \tag{49}$$

Let $\lambda := \max\{\eta, \gamma\}$. We claim that

$$\left\|\Pi_{\mathcal{U},p}(x) - \Pi_{\mathcal{U},p}(y)\right\| \le \lambda \cdot \|x - y\| \qquad \text{for } x, y \in \mathbb{R}^n,\ 1 < p < \infty, \tag{50}$$

which will complete the proof.

For $x, y \in \mathbb{R}^n$, define $x^\theta = \theta y + (1-\theta)x$ and

$$\theta_0 := \max\Big\{\theta\colon 0 \le \theta \le 1, \big\|\Pi_{U,p}(x^\theta) - \Pi_{U,p}(x)\big\| \le \lambda \cdot \|x^\theta - x\|\Big\}. \tag{51}$$

If $\theta_0 = 1$, then (50) holds. Assume to the contrary that $0 \le \theta_0 < 1$. We discuss two cases. First assume that there is $\theta_0 < \theta_1 \le 1$ such that $x^\theta \in D_p$ for $\theta_0 < \theta < \theta_1$. Then, by the continuity of $\Pi_{U,p}$ and (41), we obtain

$$\begin{aligned} &\big\|\Pi_{U,p}(x^{\theta_0}) - \Pi_{U,p}(x^{\theta_1})\big\| \\ &\quad= \lim_{\bar\theta_0 \to \theta_0^+} \lim_{\bar\theta_1 \to \theta_1^-} \big\|\Pi_{U,p}(x^{\bar\theta_0}) - \Pi_{U,p}(x^{\bar\theta_1})\big\| \\ &\quad\le \lim_{\bar\theta_0 \to \theta_0^+} \lim_{\bar\theta_1 \to \theta_1^-} \gamma \cdot \|x^{\bar\theta_0} - x^{\bar\theta_1}\| \le \lambda \cdot \|x^{\theta_0} - x^{\theta_1}\|. \end{aligned}$$

Otherwise,

$$\left\{x^\theta\colon \theta_0 < \theta < \theta_0 + \frac{1}{k}\right\} \cap \left(\bigcup_{i=1}^{n} D_{p,i}\right) \ne \emptyset \qquad \text{for } k = 1, 2, \dots .$$

Therefore an index i exists such that

$$\left\{x^\theta\colon \theta_0 < \theta < \theta_0 + \frac{1}{k}\right\} \cap D_{p,i} \ne \emptyset \qquad \text{for } k = 1, 2, \dots .$$

As a consequence, there exist $1 > \theta_1 > \theta_2 > \cdots \theta_k > \cdots \ge \theta_0$ such that $x^{\theta_k} \in D_{p,i}$ and

$$\lim_{k\to\infty} x^{\theta_k} = x^{\theta_0}.$$

By (49) and the continuity of $\Pi_{U,p}$ we derive that

$$\begin{aligned} \big\|\Pi_{U,p}(x^{\theta_0}) - \Pi_{U,p}(x^{\theta_1})\big\| &= \lim_{k\to\infty} \big\|\Pi_{U,p}(x^{\theta_k}) - \Pi_{U,p}(x^{\theta_1})\big\| \\ &\le \lim_{k\to\infty} \eta \cdot \|x^{\theta_k} - x^{\theta_1}\| \le \lambda \cdot \|x^{\theta_0} - x^{\theta_1}\|. \end{aligned}$$

So, in both cases, we have

$$\begin{aligned} &\big\|\Pi_{U,p}(x^{\theta_1}) - \Pi_{U,p}(x)\big\| \\ &\quad\le \big\|\Pi_{U,p}(x^{\theta_1}) - \Pi_{U,p}(x^{\theta_0})\big\| + \big\|\Pi_{U,p}(x^{\theta_0}) - \Pi_{U,p}(x)\big\| \\ &\quad\le \lambda \cdot \|x^{\theta_1} - x^{\theta_0}\| + \lambda \cdot \|x^{\theta_0} - x\| = \lambda \cdot \|x^{\theta_1} - x\|. \end{aligned} \tag{52}$$

By the definition of θ_0, we must have $\theta_0 \ge \theta_1$, a contradiction to the choice of θ_1. So $\theta_0 = 1$ and (50) holds. ∎

Remark (*A Second Proof of Theorem* 1). Note that our proof of Theorem 1 is self-contained. There is a shorter proof of Theorem 1 by making use of the estimate (26) of Holmes and Kripke. Let U be an m-dimensional subspace of $\mathbb{R}^n$ and let

$$D_p := \left\{x \in \mathbb{R}^n : \left(x - \Pi_{U,p}(x)\right)_i \neq 0 \qquad \text{for } 1 \leq i \leq n\right\}.$$

Define

$$J := \left\{i : z_i \neq 0 \quad \text{for some } z \in U^{\perp}\right\}. \tag{53}$$

Note that $i \in \{1, \dots, n\} \setminus J$ if and only if $e_i \in U$. Hence we get the representation

$$U = \{u \in U : u_i = 0 \qquad \text{for } i \notin J\} \oplus \{u \in \mathbb{R}^n : u_i = 0 \qquad \text{for } i \in J\}$$

and, for $x \in \mathbb{R}^n$,

$$\Pi_{U,p}(x)_i = \begin{cases} x_i & \text{for } i \notin J, \\ \left[\Pi_{U_J,p}(x_J)\right]_i & \text{for } i \in J. \end{cases} \tag{54}$$

Thus, it suffices to prove Theorem 1 for $U_J := \{u_J : u \in U\}$. However, by the definition of J, $(U_J)^{\perp} = (U^{\perp})_J$, containing a vector z such that $z_i \neq 0$ for $i \in J$. Replacing U_J by U, we may assume that for each i the subspace $U^{\perp}$ contains a vector z such that $z_i \neq 0$. Under this assumption we will show that the complement of D_p is a set of measure zero.

Let T_p be the continuous bijection of $\mathbb{R}^n$ defined by

$$\left[T_p(y)\right]_i = \operatorname{sign}(y_i)|y_i|^{1/(p-1)} \qquad \text{for } y \in \mathbb{R}^n.$$

Let $H_i := \{x \in \mathbb{R}^n : x_i = 0\}$ be a coordinate hyperplane in $\mathbb{R}^n$. By the Kolmogorov characterization of best approximations, the kernel of $\Pi_{U,p}$ (the so-called *metric complement* of U), denoted by $\ker(\Pi_{U,p})$, is exactly $T_p(U^{\perp})$.

Since there exists $z \in U^{\perp} \setminus H_i$, the dimension of the linear subspace $U^{\perp} \cap H_i$ is less than $\dim U^{\perp} = n - m$. So $T_p(U^{\perp} \cap H_i)$ is a manifold of dimension less than $(n - m)$ and $U + T_p(U^{\perp} \cap H_i)$ is a manifold of dimension less than n. Therefore, $U + T_p(U^{\perp} \cap H_i)$ has measure zero in $\mathbb{R}^n$. Note that

$$\begin{aligned} \mathbb{R}^n \setminus D_p &= \bigcup_{i=1}^{n} \left\{x \in \mathbb{R}^n : \left[x - \Pi_{U,p}(x)\right]_i = 0\right\} \\ &= \bigcup_{i=1}^{n} \left\{\left[\ker(\Pi_{U,p}) \cap H_i\right] + U\right\} \\ &= \bigcup_{i=1}^{n} \left\{\left[T_p(U^{\perp}) \cap T_p(H_i)\right] + U\right\} = \bigcup_{i=1}^{n} \left\{T_p\left(U^{\perp} \cap H_i\right) + U\right\}. \end{aligned}$$

Consequently, $\mathbb{R}^n \setminus D_p$ is a set of measure zero.

By Theorem 8, the Jacobian $\nabla_x \Pi_{U,p}$ of the Lipschitz mapping $\Pi_{U,p}$ is bounded by γ almost everywhere in $\mathbb{R}^n$. By a well-known characterization of Lipschitz continuous functions (cf. [20, Sect. 2.2.1, Chap. VI]), we obtain

$$\left\| \Pi_{U,p}(x) - \Pi_{U,p}(y) \right\| \le \gamma \cdot \|x - y\| \qquad \text{for } x, y \in \mathbb{R}^n, 1 < p < \infty. \quad \blacksquare$$

One advantage of the second proof is that it leads to an explicit estimate of an upper bound for λ_p in (26). Let J be defined as in (53) and let Q be an $\bar{n} \times \overline{m}$ matrix whose columns form an orthonormal basis of U_J. Then from the above proof and the remark after Theorem 8, we obtain

$$\left\| \Pi_{U_J,p}(x_J) - \Pi_{U_J,p}(y_J) \right\| \le \bar{\gamma} \cdot \|x - y\| \qquad \text{for } x, y \in \mathbb{R}^n, 1 < p < \infty, \tag{55}$$

where

$$\bar{\gamma} := \max_{\bar{J}} \left\{ \left\| \begin{pmatrix} I_{\bar{n}-\overline{m}} & -Q_{\bar{J}^c} Q_{\bar{J}}^{-1} \\ 0 & Q_{\bar{J}}^{-1} \end{pmatrix} \right\| : Q_{\bar{J}} \text{ is a nonsinglar } \overline{m} \times \overline{m} \text{ matrix} \right\}.$$

By (54) and (55) we obtain

$$\left\| \Pi_{U,p}(x) - \Pi_{U,p}(y) \right\| \le \bar{\gamma} \cdot \|x - y\| \qquad \text{for } x, y \in \mathbb{R}^n, 1 < p < \infty. \tag{56}$$

REFERENCES

1. N. N. Abdelmalek, Computing the strict Chebyshev solution of overdetermined linear equations, *Math. Comp.* **31** (1977), 974–983.
2. F. Deutsch, W. Li, and S.-H. Park, Characterizations of continuous and Lipschitz continuous metric selections in normed linear spaces, *J. Approx. Theory* **58** (1989), 297–314.
3. C. S. Duris and M. G. Temple, A finite step algorithm for determining the "strict" best approximation, *SIAM J. Numer. Anal.* **10** (1973), 690–699.
4. A. Egger and R. Huotari, Polya properties in $\mathbb{R}^n$, *Numer. Funct. Anal. Optim.* **13** (1992), 123–128.
5. M. Finzel, On the continuity of the strict approximation, *Rend. Circ. Mat. Palermo (2) Suppl.* **33** (1993), 301–310.
6. M. Finzel, Linear approximation in l_n^∞, *J. Approx. Theory* **76** (1994), 326–350.
7. M. Finzel, Best linear approximation in $l_1(n)$, *East J. Approx.* **2** (1996), 1–30.
8. J. Fischer, The convergence of the best discrete linear L_p approximation as $p \to 1$, *J. Approx. Theory* **39** (1983), 374–385.
9. R. Fletcher, J. A. Grant, and M. D. Hebden, Linear minimax approximation as the limit of best L_p-approximation, *SIAM J. Numer. Anal.* **11** (1974), 123–136.
10. D. Flores de Chela, Approximations in the l_∞ norm and the generalized inverse, *Linear Algebra Appl.* **42** (1982), 3–28.

11. A. J. Hoffman, Approximate solutions of systems of linear inequalities, *J. Res. Nat. Bur. Standards* **49** (1952), 263–265.
12. R. Holmes and B. Kripke, Smoothness of approximation, *Michigan Math. J.* **15** (1968), 225–248.
13. R. Huotari and W. Li, The continuity of metric projection in $l_\infty(n)$, the Pólya algorithm, the strict best approximation, and tubularity of convex sets, *J. Math. Anal. Appl.* **182** (1994), 836–856.
14. D. Jackson, Note on the median of a set of numbers, *Bull. Amer. Math. Soc.* **27** (1921), 160–164.
15. D. Landers and L. Rogge, Natural choice of L_1-approximants, *J. Approx. Theory* **33** (1981), 268–280.
16. W. Li, Best approximations in polyhedral spaces and linear programs, *in* "Approximation Theory" (G. Anastassiou, Ed.), pp. 393–400, Lecture Notes in Pure and Applied Mathematics, Vol. 138, Dekker, New York, 1992.
17. W. Li, A. J. Hoffman's theorem and metric projections in polyhedral spaces, *J. Approx. Theory* **75** (1993), 107–111.
18. W. Li, The sharp Lipschitz constants for feasible and optimal solutions of a perturbed linear program, *Linear Algebra Appl.* **187** (1993), 15–40.
19. B. S. Mityagin, Extremal points of a family of convex functions, *Siberian Math. J.* **6** (1965), 556–563.
20. E. M. Stein, Singular integrals and differentiability properties of functions, *in* "Princeton Math. Ser.," Vol. 30, Princeton Univ. Press, Princeton, NJ, 1970.