Analysis and Applications, Vol. 19, No. 1 (2021) 21–42
© World Scientific Publishing Company
DOI: 10.1142/S0219530520400059

World Scientific
www.worldscientific.com

Minimum norm interpolation in the $\ell_1(\mathbb{N})$ space

Raymond Cheng* and Yuesheng Xu†
Department of Mathematics and Statistics
Old Dominion University, Norfolk 23529, Virginia, USA
*rcheng@odu.edu
†ylxu@odu.edu

Received 20 November 2019
Accepted 11 March 2020
Published 17 July 2020

We consider the minimum norm interpolation problem in the $\ell_1(\mathbb{N})$ space, aiming at constructing a sparse interpolation solution. The original problem is reformulated in the pre-dual space, thereby inducing a norm in a related finite-dimensional Euclidean space. The dual problem is then transformed into a linear programming problem, which can be solved by existing methods. With that done, the original interpolation problem is reduced by solving an elementary finite-dimensional linear algebra equation. A specific example is presented to illustrate the proposed method, in which a sparse solution in the $\ell_1(\mathbb{N})$ space is compared to the dense solution in the $\ell_2(\mathbb{N})$ space. This example shows that a solution of the minimum norm interpolation problem in the $\ell_1(\mathbb{N})$ space is indeed sparse, while that of the minimum norm interpolation problem in the $\ell_2(\mathbb{N})$ space is not.

Keywords: Minimum norm interpolation; duality; linear programming; sparsity.

Mathematics Subject Classification 2020: 41A05, 46B45

1. Introduction

Minimum norm interpolation in a Hilbert space is a classical research topic [8, 13]. In particular, minimum norm interpolation in the $\ell_2(\mathbb{N})$ space produces good results in the sense of approximation. However, due to the roundness of the unit ball of a Hilbert space such as $\ell_2(\mathbb{N})$, the resulting minimum norm interpolation solution is normally represented by a dense vector, in the sense that a majority of its components are nonzero. Dense vectors are less computationally efficient for high dimensional problems. Thus for potential use in treating big data sets, we prefer a sparse vector, in the sense that a majority of its components are zero, for representing a minimum norm interpolation solution. For this purpose, we consider minimum norm interpolation in the $\ell_1(\mathbb{N})$ space.

The choice of the ℓ_1 space is also motivated by recent exciting progress in signal processing and machine learning. Compressed sensing [4, 9] based on the ℓ_1 norm

†Corresponding author.

has attracted much attention in recent application areas such as signal processing and image processing. The use of the ℓ_1 norm in compressed sensing gives rise to desirable sparsity in the resulting solutions [1]. Seeking sparse solutions in machine learning leads to the introduction of reproducing kernel Banach spaces [22]. For recent developments in reproducing kernel Banach spaces, the readers are referred to [15, 16, 21]. Representer theorems for the solution of ill-posed linear inverse problems are provided in [18], comparing the effects of ℓ_1 and ℓ_2 regularization. In [17] a representer theorem is obtained in the general Banach space setting. Such representer theorems enable a dramatic reduction in the dimension of the space from which a data fitting solution needs to be sought.

The contribution of this paper is to furnish a method for solving the minimum norm interpolation problem in ℓ_1 that determines a sparse solution. In this method, we transform the minimum norm interpolation problem into two related finite-dimensional problems, for which established solution methods exist. We first reformulate the proposed minimum norm interpolation problem using a duality argument. This process introduces a norm in a related finite-dimensional Euclidian space. The associated dual extremal problem then takes the form of a basic linear programming problem, namely, optimizing a linear function on a convex polytope. There is a substantial literature on optimizing a linear function on a convex polytope (see, for example, [12, 19]). Finally, a solution of the linear programming problem enables the original interpolation problem to be reduced to an elementary equation in finite-dimensional linear algebra. Again, this equation yields to well established methods.

We organize this paper in five sections. In Sec. 2, we describe the minimum norm interpolation problem in the $\ell_1(\mathbb{N})$ space, and show that it has a solution. We then reformulate it via a Banach space duality argument in Sec. 3. We introduce in Sec. 4 a norm in a related finite-dimensional Euclidean space, and further reformulate the dual problem as an equivalent linear programming problem. In Sec. 5, we leverage the solution of the linear programming problem into a solution of the original interpolation problem. Finally, in Sec. 6, we present an example of solving the problem completely by using the proposed approach, and compare it to the Hilbert space approach.

2. The ℓ_1 Interpolation Problem

In this section, we present the minimum ℓ_1 norm interpolation problem, which is the principal subject of this paper. To frame it properly we review the classical Banach spaces ℓ_1 and c_0. We then argue that the interpolation problem has a solution under natural conditions, to close the section.

We now introduce the main problem under investigation. By $\ell_1 := \ell_1(\mathbb{N})$ we mean the Banach space of real sequences $\mathbf{x} := (x_1, x_2, \ldots)$ such that

$$\|\mathbf{x}\|_1 := \sum_{k=1}^{\infty} |x_k| < \infty.$$

The space c_0 is the set of real sequences that are convergent to zero. Thus, for $\mathbf{a} \in c_0$, we can write

$$\mathbf{a} := (a_1, a_2, \ldots)$$

and by definition

$$\lim_{k \to \infty} a_k = 0.$$

The set c_0 is clearly a linear space, and in fact it is a Banach space under the supremum norm. That is, for $\mathbf{a} := (a_1, a_2, \ldots) \in c_0$, we define

$$\|\mathbf{a}\|_\infty := \sup\{|a_k| : k \in \mathbb{N}\}.$$

The scalar field is $\mathbb{R}$. We observe that for an $\mathbf{a} \in c_0$, it holds that $\|\mathbf{a}\|_\infty < +\infty$. For any $\mathbf{x} \in \ell_1$ and $\mathbf{a} \in c_0$, let us write

$$\langle \mathbf{a}, \mathbf{x} \rangle := \sum_{k=1}^{\infty} a_k x_k.$$

Clearly, we have that

$$|\langle \mathbf{a}, \mathbf{x} \rangle| \leq \sum_{k=1}^{\infty} |a_k x_k| \leq \|\mathbf{a}\|_\infty \|x\|_1 < +\infty.$$

In other words, for any $\mathbf{x} \in \ell_1$ and $\mathbf{a} \in c_0$, the quantity $\langle \mathbf{a}, \mathbf{x} \rangle$ is well-defined.

We recall the notion of a continuous (or bounded) linear functional on c_0. A continuous linear functional on a Banach space $\mathscr{X}$ is a linear function $\lambda : \mathscr{X} \longmapsto \mathbb{R}$ that is continuous with respect to the metric topologies on $\mathscr{X}$ and $\mathbb{R}$. The dual of a Banach space $\mathscr{X}$ is the set of continuous linear functionals on that space. The dual space, given the symbol $\mathscr{X}^*$, is a Banach space in its own right, endowed with the norm

$$\|\lambda\|_{\mathscr{X}^*} = \sup\left\{\frac{|\lambda(\mathbf{x})|}{\|\mathbf{x}\|_{\mathscr{X}}} : \mathbf{x} \neq 0\right\}.$$

For any $\mathbf{x} \in \ell_1$, the mapping

$$\mathbf{a} \mapsto \langle \mathbf{a}, \mathbf{x} \rangle \tag{2.1}$$

induces a bounded linear functional on c_0. More will be said about this situation in the next section.

We now describe the minimum ℓ_1 norm interpolation problem. Let $S := \{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m\}$ be a set of given sequences from c_0, and let $\{y_1, y_2, \ldots, y_m\}$ be a set of real numbers. Consider the problem of finding $\mathbf{x} \in \ell_1$ such that the infimum

$$m_S := \inf\{\|\mathbf{x}\|_1 : \langle \mathbf{a}_j, \mathbf{x} \rangle = y_j, \text{ for all } 1 \leq j \leq m\} \tag{2.2}$$

is achieved. Of course the existence of such a vector needs to be established, and we address this matter below.

We first consider the linear independence assumption on the vectors of S. To this end, we introduce an mth order semi-infinite matrix $\mathbf{A}$ whose rows are the members of S. The system of equations

$$\langle \mathbf{a}_j, \mathbf{x}\rangle = y_j, \quad 1 \le j \le m \tag{2.3}$$

can be equivalently expressed in a matrix form as

$$\mathbf{A}\mathbf{x} = \mathbf{y}, \tag{2.4}$$

where $\mathbf{x}$ is an infinite column vector representing an element of ℓ_1, and $\mathbf{y}$ is an m-dimensional column vector. By elementary matrix algebra, there exist an $m \times m$ permutation matrix $\mathbf{P}$ and an $m \times m$ lower triangular matrix $\mathbf{L}$ such that

$$\mathbf{PA} = \mathbf{LU},$$

and $\mathbf{U}$ is in reduced echelon form. We may insist that $\mathbf{L}$ bas unit entries along the main diagonal, since it is a product of elementary matrices and a permutation matrix. Then the system (2.4) can be transformed to

$$\mathbf{U}\mathbf{x} = \mathbf{L}^{-1}\mathbf{P}\mathbf{y}.$$

If the vectors of S are linearly dependent, then the system could be inconsistent (depending on the numbers y_j), in which case there is no solution. In this situation, there would be rows of all zeros in $\mathbf{U}$, corresponding to nonzero entries in the column matrix $\mathbf{L}^{-1}\mathbf{P}\mathbf{y}$. On the other hand, if the system is consistent, then by discarding any rows of zeros in $\mathbf{U}$, as well as the corresponding zero components of the column vector $\mathbf{L}^{-1}\mathbf{P}\mathbf{y}$, we obtain an equivalent system where the rows of the truncated $\mathbf{U}$ are linearly independent. We may therefore assume at the outset that any superfluous vectors from S have been discarded, and thus S is linearly independent and the system is consistent. This ensures that the infimum in (2.2) is over a nonempty set. To rule out further trivialities, let us also assume that the infimum in (2.2) is positive. This is to say that $\mathbf{y}$ is not the zero vector.

For convenience, let $\mathscr{S}$ denote the collection of vectors $\mathbf{x}$ in ℓ_1 satisfying the system (2.3). Thus the extremal problem (2.2) could be written as

$$\inf\{\|\mathbf{x}\|_1 : \mathbf{x} \in \mathscr{S}\}.$$

We write $\mathscr{A}$ for the closed linear span in c_0 of the vectors of S. Let us show that under the conditions described earlier, the infimum in (2.2) must be attained. A proof of this result could be fashioned using the Banach–Alaoglu theorem. However, since c_0 is separable, the following elementary argument is made possible. It essentially re-proves the Banach–Alaoglu theorem in the special case that the pre-dual is separable.

Proposition 2.1. *Let m be a positive integer. Suppose that $S := \{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m\}$ is a linearly independent set of sequences from c_0, and $\{y_1, y_2, \ldots, y_m\}$ is a set of real numbers, not all zero. If the system* (2.3) *is consistent, then there exists $\mathbf{x}_0 \in \mathscr{S}$ satisfying*

$$\|\mathbf{x}_0\|_1 = m_S, \tag{2.5}$$

where m_S is defined by (2.2).

Proof. We prove the claim by a "minimizing sequence" argument. Suppose that $\{\mathbf{x}^{(n)}\}_{n=1}^{\infty}$ is a sequence of vectors in ℓ_1 satisfying the system (2.3) such that

$$\lim_{n\to\infty} \|\mathbf{x}^{(n)}\|_1 = m_S. \tag{2.6}$$

The proof will be completed by first constructing the limit $\mathbf{x}_0$ of a subsequence of the given minimizing sequence, showing that it is in ℓ_1, then verifying it satisfies the system (2.3), and finally proving that $\mathbf{x}_0$ satisfies (2.5).

We first construct the element $\mathbf{x}_0$. By (2.6) the sequence $\{\|\mathbf{x}^{(n)}\|_1\}_{n=1}^{\infty}$ is convergent, hence bounded, and let C be its supremum. We use the notation $\mathbf{x}^{(n)} := (x_1^{(n)}, x_2^{(n)}, \ldots)$ for the components of $\mathbf{x}^{(n)}$. Then the real sequence $(x_1^{(n)})_{n=1}^{\infty}$ is bounded; by the Bolzano-Weierstrass theorem, there is a convergent subsequence. Let this subsequence arise from the indices $(n_{1,j})_{j=1}^{\infty}$, and let the x_1 be the limit of $x_1^{(n_{1,j})}$ as j tends to infinity. Having defined $x_1, x_2, \ldots, x_k$, and having selected the indices $(n_{k,j})_{j=1}^{\infty}$ for a subsequence of $\{\mathbf{x}^{(n)}\}_{n=1}^{\infty}$, we observe that $(x_{k+1}^{(n_{k,j})})_{j=1}^{\infty}$ is a bounded real sequence. Hence this subsequence has a further subsequence, with indices $(n_{k+1,j})_{j=1}^{\infty}$, convergent to a limit x_{k+1}. In this manner, a vector $\mathbf{x}_0 := (x_1, x_2, \ldots)$ is specified.

We next verify that $\mathbf{x}_0$ is in ℓ_1. To this end, for any $N \in \mathbb{N}$, we can choose $\nu(N)$ sufficiently large such that $n_{N,\nu(N)} > n_{N-1,\nu(N-1)}$, for all $N \geq 2$, and

$$|x_j| \leq |x_j^{(n_{N,\nu(N)})}| + \frac{1}{2^j}, \quad \text{for all } 1 \leq j \leq N.$$

Summing both sides of the above inequality leads to

$$\sum_{j=1}^{N} |x_j| \leq \|\mathbf{x}^{(n_{N,\nu(N)})}\|_1 + \sum_{j=1}^{N} \frac{1}{2^j}.$$

This implies that

$$\|\mathbf{x}_0\|_1 \leq C + 1. \tag{2.7}$$

That is, $\mathbf{x}_0 \in \ell_1$.

We now show that $\mathbf{x}_0$ satisfies the system (2.3). To accomplish this, we establish the weak* convergence of $\mathbf{x}^{(n_{N,\nu(N)})}$ to $\mathbf{x}$. That is, for any $\mathbf{a} \in c_0$

$$\lim_{N\to\infty} (\mathbf{a}, \mathbf{x}^{(n_{N,\nu(N)})}) = (\mathbf{a}, \mathbf{x}_0). \tag{2.8}$$

For this purpose, we let $\mathbf{e}_k$ denote the vector in ℓ_1 whose kth component is equal to 1 and all other components are zero. By the construction of $\mathbf{x}_0$, for each k we have that

$$\lim_{N\to\infty} (\mathbf{e}_k, \mathbf{x}^{(n_{N,\nu(N)})}) = x_k = (\mathbf{e}_k, \mathbf{x}_0).$$

The continuity of $(\cdot,\cdot)$ then allows for

$$\lim_{N\to\infty} \left\langle \sum_{k=1}^{K} a_k \mathbf{e}_k, \mathbf{x}^{(n_{N,\nu(N)})} \right\rangle = \left\langle \sum_{k=1}^{K} a_k \mathbf{e}_k, \mathbf{x}_0 \right\rangle$$

for any linear combination $\sum_{k=1}^{K} a_k \mathbf{e}_k$. Now, suppose that $\lim_{K\to\infty} \sum_{k=1}^{K} a_k \mathbf{e}_k = \mathbf{a}$ in the norm topology of c_0. Let $\epsilon > 0$ be chosen. There exists an index K sufficiently large such that

$$\left\| \mathbf{a} - \sum_{k=0}^{K} a_k \mathbf{e}_k \right\|_{\infty} \leq \frac{\epsilon}{2(2C+1)}. \tag{2.9}$$

With K fixed, there exists an index n sufficiently large such that whenever $N \geq n$,

$$\left| \left\langle \sum_{k=1}^{K} a_k \mathbf{e}_k, \mathbf{x}^{(n_N,\nu(N))} - \mathbf{x}_0 \right\rangle \right| \leq \frac{\epsilon}{2}.$$

Then, we observe that

$$\begin{aligned} &|(\mathbf{a}, \mathbf{x}^{(n_N,\nu(N))} - \mathbf{x}_0)| \\ &\leq \left| \left\langle \mathbf{a} - \sum_{k=1}^{K} a_k \mathbf{e}_k, \mathbf{x}^{(n_N,\nu(N))} - \mathbf{x}_0 \right\rangle \right| + \left| \left\langle \sum_{k=1}^{K} a_k \mathbf{e}_k, \mathbf{x}^{(n_N,\nu(N))} - \mathbf{x}_0 \right\rangle \right|. \end{aligned}$$

The second term on the right-hand side of the inequality above is bounded by $\frac{\epsilon}{2}$ and using (2.7) the first term on the right-hand side is bounded by $\left\| \mathbf{a} - \sum_{k=1}^{K} a_k \mathbf{e}_k \right\|_{\infty} (2C+1)$, which is also bounded by $\frac{\epsilon}{2}$ by employing (2.9). Therefore, we have established (2.8). By choosing $\mathbf{a} := \mathbf{a}_j$ in (2.8) and noticing that vectors $\mathbf{x}^{(n_N,\nu(N))}$ for all N satisfy the system (2.3), we see that $\mathbf{x}_0$ must satisfy the system (2.3).

Finally, we establish that $\mathbf{x}_0$ satisfies (2.5). To this end, we consider vector $\mathbf{b}$ of the form

$$b_j = \begin{cases} \operatorname{sign}(x_j), & 1 \leq j \leq M, \\ 0, & j > M, \end{cases}$$

which have the unit norm in c_0. For any $\epsilon > 0$ we can find M sufficiently large such that

$$\begin{aligned} \|\mathbf{x}_0\|_1 - \epsilon \leq \sum_{j=1} |x_j| &= |(\mathbf{b}, \mathbf{x}_0)| \\ &= \lim_{n\to\infty} |(\mathbf{b}, \mathbf{x}^{(n_N,\nu(N))})| \leq \liminf_{n\to\infty} \|\mathbf{x}^{(n_N,\nu(N))}\|_1 \|\mathbf{b}\|_{c_0}. \end{aligned}$$

Since $\|\mathbf{b}\|_{c_0} = 1$, this demonstrates that

$$\|\mathbf{x}_0\|_1 \leq \lim_{n\to\infty} \|\mathbf{x}^{(n_N,\nu(N))}\|_1 + \epsilon, \quad \text{for every } \epsilon > 0.$$

Equality must hold (since $\mathbf{x}_0 \in \mathscr{S}$). Furthermore, noticing $\mathbf{x}^{(n_N,\nu(N))}$ is a subsequence of the minimizing sequence, we observe that

$$\lim_{n\to\infty} \|\mathbf{x}^{(n_N,\nu(N))}\|_1 = m_S.$$

Thus, the infimum in (2.5) is attained by $\mathbf{x}_0$. □

Proposition 2.1 ensures that the main problem expressed in (2.2) has a solution under very broad conditions. We emphasize that the infimum of (2.2) need not be *uniquely* attained, since the ball in ℓ_1 fails to be strictly convex.

We close this section by noting that the infimum problem (2.2) is closely related to the "regularized" problem of minimizing the quantity

$$\sum_{j=1}^{m} |\langle \mathbf{a}_j, \mathbf{x}\rangle - y_j| + \Phi(\|\mathbf{x}\|_1), \tag{2.10}$$

where Φ is a monotone increasing function on the positive real axis, as $\mathbf{x}$ varies through ℓ^1. That is, we are willing to trade off the exact equality of each $\langle \mathbf{a}_j, \mathbf{x}\rangle = y_j$, in return for keeping the down the value of the norm $\|\mathbf{x}\|_1$. Indeed, this second term in (2.10) is intended to penalize "overfitting" of the data. Regularization problems for machine learning in reproducing kernel Hilbert spaces have been well-studied in [7, 10]. For learning a matrix by regularization, see [2]. Moreover, for regularization problems in functional reproducing kernel Hilbert spaces, see [20], and in reproducing kernel Banach spaces, see [15, 21, 22]. The regularized problem related to (2.10) is studied from a function-theoretic approach in a forthcoming paper.

3. Dual Extremal Problem

In this section, we reformulate the infimum problem (2.2) as a dual extremal problem. This duality argument is a well established tool in convex analysis. In its general form can be found in any standard text in functional analysis, for example, [6]. A particular application of duality to machine learning appears in [2, Sec. 3], to solve the problem of learning a matrix based on a set of linear measurements. It results in a significant reduction in the number of free parameters, and hence on the computational burden. We shall see that a similar reduction is enabled in the present paper.

The duality argument applied here is made possible by the following relationship between the spaces c_0 and ℓ_1. For a derivation of this well-known result, see [14, pp. 73–74].

Proposition 3.1. *The dual space of c_0 is ℓ_1; that is, $c_0^* = \ell_1$.*

Let us note that the set of vectors $\mathscr{S}$ over which the infimum is being taken is a certain hyperplane in ℓ_1. To this end, we define a subspace of ℓ_1 by letting

$$\mathscr{M} := \{\mathbf{z} \in \ell_1 : \langle \mathbf{a}_j, \mathbf{z}\rangle = 0, \text{ for all } 1 \le j \le m\}. \tag{3.1}$$

Lemma 3.2. *If $\mathbf{x}' \in \mathscr{S}$ is any particular solution to the system* (2.3), *then*

$$\inf\{\|\mathbf{x}' + \mathbf{z}\|_1 : \mathbf{z} \in \mathscr{M}\} = m_S. \tag{3.2}$$

This lemma says that $\mathscr{S} = \mathbf{x}' + \mathscr{M}$. That is, the set $\mathscr{S}$ over which the infimum (2.2) is taken is a translation of the subspace $\mathscr{M}$ of ℓ_1. The proof is elementary and hence omitted.

The following proposition enables us to reformulate the extremal problem by means of classical Banach space duality. The result follows from a basic theorem from functional analysis, namely, if $\mathscr{A}$ is a subspace of a Banach space $\mathscr{X}$, then the dual of $\mathscr{A}$ is isometrically isomorphic to the quotient space $\mathscr{X}^*/\mathscr{A}^\perp$, where

$$\mathscr{A}^\perp := \{\lambda \in \mathscr{X}^* : \lambda(\mathbf{a}) = 0, \text{ for all } \mathbf{a} \in \mathscr{A}\}$$

is the annihilator of $\mathscr{A}$. For this we refer the reader to [6, Theorem 2.3].

Proposition 3.3. *If* $\mathbf{x}' \in \mathscr{S}$, *then*

$$\inf\{\|\mathbf{x}' + \mathbf{z}\|_1 : \mathbf{z} \in \mathscr{M}\} = \sup_{(c_1,c_2,\ldots,c_m)\in\mathbb{R}^m} \frac{\sum_{j=1}^m c_j y_j}{\|\sum_{j=1}^m c_j \mathbf{a}_j\|_\infty}. \tag{3.3}$$

Proof. Here is an elementary proof of [6, Theorem 2.3] for the present setting, where $\mathscr{X} = c_0$, and $\mathscr{A}$ is the closed linear span in c_0 of $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$. (Note that $\mathscr{M} = \mathscr{A}^\perp$ here.) Let us regard the given vector $\mathbf{x}' \in \ell_1$ as a bounded linear functional restricted to $\mathscr{A}$. By the Hahn Banach Theorem, there exists a bounded linear functional on all of c_0, represented by some $\mathbf{x} \in \ell_1$, such that $\mathbf{x}$ agrees with $\mathbf{x}'$ on $\mathscr{A}$, and

$$\|\mathbf{x}\|_1 = \sup_{(c_1,c_2,\ldots,c_m)\in\mathbb{R}^m} \frac{(\sum_{j=1}^m c_j \mathbf{a}_j, \mathbf{x}')}{\|\sum_{j=1}^m c_j \mathbf{a}_j\|_\infty}.$$

That is, the norm of $\mathbf{x}$ equals the norm of the restriction of $\mathbf{x}'$ to $\mathscr{A}$. Direct computation on the right-hand side of the above equation leads to

$$\|\mathbf{x}\|_1 = \sup_{(c_1,c_2,\ldots,c_m)\in\mathbb{R}^m} \frac{\sum_{j=1}^m c_j y_j}{\|\sum_{j=1}^m c_j \mathbf{a}_j\|_\infty}. \tag{3.4}$$

Since $\mathbf{x}$ and $\mathbf{x}'$ are equal when restricted to $\mathscr{A}$, their difference $\mathbf{z} := \mathbf{x} - \mathbf{x}'$ belongs to $\mathscr{M}$. Thus, we have that

$$\inf\{\|\mathbf{x}' + \mathbf{z}\|_1 : \mathbf{z} \in \mathscr{M}\} = \|\mathbf{x}\|_1.$$

This combined with Eq. (3.4) leads the desired result (3.3). □

According to Lemma 3.2 and Proposition 3.3, solving the original minimum norm interpolation problem is equivalent by solving the dual extremal problem

$$\sup_{(c_1,c_2,\ldots,c_m)\in\mathbb{R}^m} \frac{\sum_{j=1}^m c_j y_j}{\|\sum_{j=1}^m c_j \mathbf{a}_j\|_\infty}. \tag{3.5}$$

Problem (3.5) has only finitely many real parameters, and therefore this step is a beneficial reduction.

We next consider the existence of a solution of the dual extremal problem (3.5). To this end, we identify a norm that arises naturally from (3.5). The following lemma can be verified by inspection.

Lemma 3.4. *Let m be a positive integer. If $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ are linearly independent vectors in c_0, then the mapping*

$$\mathbf{c} := (c_1, c_2, \ldots, c_m) \mapsto \left\| \sum_{j=1}^{m} c_j \mathbf{a}_j \right\|_\infty \tag{3.6}$$

is a norm on $\mathbb{R}^m$.

In the remaining part of this paper, we shall always assume that $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ are *fixed* linearly independent vectors in c_0 without further mentioning. Let us give the associated norm the name

$$\|\mathbf{c}\|_\star := \left\| \sum_{j=1}^{m} c_j \mathbf{a}_j \right\|_\infty$$

We prove in the next proposition the existence of a solution of the extremal problem (3.5).

Proposition 3.5. *The supremum in* (3.5) *is attained by some choice of* $(c_1, c_2, \ldots, c_m) \in \mathbb{R}^m$.

Proof. The mapping

$$(c_1, c_2, \ldots, c_m) \mapsto \sum_{j=1}^{m} c_j y_j$$

is a continuous function from $\mathbb{R}^m$ to $\mathbb{R}$. In (3.5) we are taking an extreme value of this function over a compact set, namely the unit sphere in $\mathbb{R}^m$ in the $\|\cdot\|_\star$ norm (recall that all norms on $\mathbb{R}^m$ give rise to equivalent topologies, and hence this set is compact under both the Euclidean topology and the $\|\cdot\|_\star$ topology). Thus the supremum of this function is attained. □

We emphasize that the extremal vector $\mathbf{c}$ for (3.5) need not be unique.

4. A Linear Programming Problem

The dual extremal problem described in the last section turns out to be equivalent to a linear programming problem. This section is devoted to establishing this equivalence.

We first recall some necessary notions from convex analysis. A *convex polytope* in $\mathbb{R}^m$ is a bounded region of $\mathbb{R}^m$ that is the intersection of finitely many halfspaces. It is equal to the convex hull of its vertices. An m-dimensional polytope is bounded by finitely many $(m-1)$-dimensional *facets*, each of which is a polytope in a lower dimensional space. Thus the notion of polytope generalizes that of a polyhedron to arbitrarily many finite dimensions. (Our source on this subject is [11].)

Presently we will see that the unit sphere in $\mathbb{R}^m$ under the $\|\cdot\|_*$ norm must be the surface of a convex polytope. This reduces the dual extremal problem to one of standard linear programming.

Proposition 4.1. *The closed unit ball B_* in $\mathbb{R}^m$ under the $\|\cdot\|_*$ norm is a convex polytope.*

Proof. We show this result by proving that B_* is the intersection of finitely many halfspaces. For each $k = 1, 2, \ldots$, let U_k denote the region in $\mathbb{R}^m$ given by

$$U_k := \left\{ \mathbf{x} \in \mathbb{R}^m : -1 \leq \sum_{j=1}^{m} x_j a_{j,k} \leq 1 \right\}. \tag{4.1}$$

Each such region is the gap enclosed between two hyperplanes. Then the closed unit ball B_* in $\mathbb{R}^m$ under $\|\cdot\|_*$ is given by

$$B_* := \bigcap_{k=1}^{\infty} U_k.$$

We claim that in fact B_* is the intersection of *finitely* many of the regions U_k. To see this, we consider again the $m \times \infty$ matrix $\mathbf{A}$ from (2.4) with entries $[a_{j,k}]_{1\leq j\leq m,\ k\geq 1}$. We denote by A^k, $k \in \mathbb{N}$, the columns of $\mathbf{A}$. By assumption the m rows are linearly independent vectors $\mathbf{a}_j$, $j = 1, 2, \ldots, m$, in c_0. Hence, there exist m linearly independent columns A^k of $\mathbf{A}$, $k \in \mathbb{N}_m := \{n_1, n_2, \ldots, n_m\} \subset \mathbb{N}$, that span the space $\mathbb{R}^m$.

With that noted, it must be that

$$B' := \bigcap_{k\in\mathbb{N}_m} U_k$$

is a bounded subset of $\mathbb{R}^m$. For if not, then by symmetry of the regions U_k and the definition of B', it must contain a line

$$L := \{\alpha\mathbf{w} \in \mathbb{R}^m : \alpha \in \mathbb{R}\}, \tag{4.2}$$

where $\mathbf{w}$ is some fixed nonzero vector in $\mathbb{R}^m$. To see why this is the case, we point out that B' must be convex, being the intersection of halfspaces; furthermore, it is symmetric about the origin, since that is true of each pair of bounding hyperplanes. Thus for any point lying in B', the entire segment connecting the point and its reflection about the origin must be contained in B' as well. Next, unboundeness would imply the existence of points $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \ldots$ belonging to B' with $\|\mathbf{w}_k\|_{\mathbb{R}^m} > k$ (where $\|\cdot\|_{\mathbb{R}^m}$ is the usual Euclidean norm). Then by compactness of the unit sphere of $\mathbb{R}^m$ there must be some subsequence of points $\mathbf{w}_{k_n}/\|\mathbf{w}_{k_n}\|_{\mathbb{R}^m}$ converging to a point $\mathbf{w}$. The closedness of B' then ensures that the entire line L defined by (4.2) is contained in B'. Further from the definition of B', the line L must be parallel to all of the hyperplanes bounding U_k, $k \in \mathbb{N}_m$. This is equivalent to saying

$$\sum_{j=1}^{m} w_j a_{j,k} = 0, \quad \text{for all } k \in \mathbb{N}_m.$$

But since these m columns of $\mathbf{A}$ span $\mathbb{R}^m$, this forces $\mathbf{w} = \mathbf{0}$, a contradiction. This contradiction rules out the possibility that B' is unbounded.

Let $\rho > 0$ be sufficiently large that B' is contained in the (Euclidean) ball of radius ρ. Now, the Euclidean distance between the two hyperplanes bounding U_k is given by

$$d_k := \frac{2}{\left(\sum_{j=1}^{m} a_{j,k}^2\right)^{\frac{1}{2}}}.$$

Since each sequence $\mathbf{a}_j$ converges to zero for all $1 \leq j \leq m$, it follows that d_k diverges to infinity as the index k increases without bound. That is, there is an index k_0 sufficiently large that

$$\frac{1}{2} d_k > \rho, \quad \text{whenever } k \geq k_0. \tag{4.3}$$

For such k, we have $B' \subseteq U_k$, and these U_k contribute nothing to the intersection defining B'. Because $B_\star \subseteq B'$, it must be that $B_\star$ can be expressed as the intersection of only finitely many of the U_k.

Finally, the convexity of $B_\star$ follows from it being the intersection of halfspaces in $\mathbb{R}^m$, which are themselves convex. □

An upper bound for the number of regions U_k contributing to the determination of B is k_0. Therefore an upper bound for the number of faces of the ball $B_\star$ is $2k_0$. Determining the vertices of $B_\star$ given its bounding hyperplanes is the "Vertex Enumeration Problem." The computational complexity of this problem is the subject of ongoing research in computer science and graph theory. For example, see [3].

We can arrive at a very crude bound for the number of vertices of $B_\star$ in the following manner. Let $2J$ points in $\mathbb{R}^m$ be given, where $J \geq m$. It takes m points to determine a hyperplane in $\mathbf{R}$. This is because the system

$$\sum_{k=1}^{m} \beta_{j,k} x_k = c, \quad \text{for all } 1 \leq j \leq m$$

where $c = 0$ or $c = 1$, will determine at most one solution $(x_1, x_2, \ldots, x_m)$, given the m points $(\beta_{j,1}, \beta_{j,2}, \ldots, \beta_{j,m}) \in \mathbf{R}^m$, for $1 \leq j \leq m$. (This is analogous to saying that three points determine at most one plane in $\mathbf{R}^3$.) Therefore the set of $2J$ given points can give rise to at most

$$N := \binom{2J}{m} = \frac{(2J)!}{(m!)(2J-m)!}$$

faces. The ball $B_\star$ is dual (in the graph-theoretic sense) to a polytope arising in this fashion. Accordingly, the number of vertices of $B_\star$ cannot exceed N.

In any case, the extreme values in (3.3) must be attained somewhere on the finitely many vertices of the convex polytope $B_\star$. This effectively reduces the dual extremal problem to one in linear programming. To see this, we begin with an obvious observation.

Lemma 4.2. *Let d be a positive integer. Suppose that f is a linear function of d real variables, i.e.,*

$$f(x_1, x_2, \ldots, x_d) := c_0 + \sum_{j=1}^{d} c_j x_j,$$

for some $c_0, c_1, \ldots, c_d \in \mathbb{R}$. If L is any line segment in $\mathbb{R}^d$, then f achieves its maximum over the points of L at an endpoint of L.

Applied repeatedly, this gives rise to an important fact about linear optimization.

Proposition 4.3. *Let m be a positive integer. If f is a linear function defined on a polytope $\Delta \in \mathbb{R}^m$, then f attains its maximum value at a vertex of Δ.*

Proof. Suppose that L is any line segment passing through Δ. By Lemma 4.2, the maximum value of f along L must be attained at an endpoint of L. This shows that the maximum of f over all of the polytope Δ cannot be achieved at an interior point of Δ.

The boundary of Δ is made up of $(m-1)$-dimensional facets. For each such facet T, Lemma 4.2 again shows that the maximum of f over T must be attained at a boundary point of T. Continuing on in this fashion, we see that the maximum of f along all of Δ must be achieved at an edge point of Δ. Apply Lemma 4.2 one more time to conclude that such a maximum occurs at an endpoint of an edge. These endpoints are the vertices of Δ. □

With the subsets U_k defined as in (4.1), and the index k_0 given by (4.3), we have therefore reformulated the dual extremal problem in the following terms.

Theorem 4.4. *The dual extremal problem* (3.5) *is equivalent to maximizing the linear function*

$$f(x_1, x_2, \ldots, x_m) := \sum_{j=1}^{m} x_j y_j$$

where $(x_1, x_2, \ldots, x_m)$ varies over the polytope

$$\Delta := \bigcap_{k=1}^{k_0} U_k.$$

Furthermore, a solution of the dual extremal problem (3.5) *is attained at one of the vertices of the polytope Δ.*

Proof. Clearly, the dual extremal problem (3.5) is equivalent to

$$\sup\left\{\sum_{j=1}^{m} c_j y_j : \mathbf{c} := (c_1, c_2, \ldots, c_m) \in \mathbb{R}^m, \|\mathbf{c}\|_* = 1\right\}. \tag{4.4}$$

By Proposition 4.3, a solution of the extremal problem (4.4) is one of the vertices of the unit ball $\{\mathbf{c} \in \mathbb{R}^m : \|\mathbf{c}\|_* \leq 1\}$, which is identified as the polytope Δ by Proposition 4.1. □

We have therefore shown that the dual extremal problem is equivalent to a standard problem in linear programming. The complexity of this linear programming problem depends on the parameter k_0 associated with the given set of vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ in c_0.

5. Solution to the ℓ_1 Interpolation Problem

From here our final objective is to leverage the solution to the dual extremal problem into a solution of the original problem (2.2). Thus far, we have found a solution to the dual extremal problem, and calculated the value of the infimum m_S in (2.2). It remains to identify vectors $\mathbf{x} \in \ell_1$ for which this extreme value arises.

Our strategy will be to use the concept of norming functional to identify candidate vectors in ℓ_1 for the solution, based on a solution to the dual problem. These candidates for the solution will turn out to constitute a finite-dimensional convex subset of the sphere in ℓ_1 with radius m_S. To find the actual solutions, it remains to re-impose the linear system (2.3). When this is done, we are left with a finite-dimensional linear algebra equation, which can be solved with well-known techniques. Solving this linear algebra equation leads to the solution of original problem (2.2).

We begin by giving the name $\mathbf{a}' := \sum_{j=1}^{m} c_j' \mathbf{a}_j$ to a vector in the sequence space c_0 for which the dual extremal problem is attained:

$$\frac{\sum_{j=1}^{m} c_j' y_j}{\|\sum_{j=1}^{m} c_j' \mathbf{a}_j\|_\infty} = \sup_{(c_1,c_2,\ldots,c_m)\in\mathbb{R}^m} \frac{\sum_{j=1}^{m} c_j y_j}{\|\sum_{j=1}^{m} c_j \mathbf{a}_j\|_\infty}.$$

We will now utilize the notion of a norming functional for a vector in a Banach space. Given a nonzero vector $\mathbf{x} \in \mathscr{X}$, a *norming functional* for $\mathbf{x}$ is a bounded linear functional $\lambda \in \mathscr{X}^*$ satisfying $\|\lambda\|_{\mathscr{X}^*} = 1$ and

$$\lambda(\mathbf{x}) = \|\mathbf{x}\|_{\mathscr{X}}.$$

The existence of a norming functional for any nonzero vector is assured by the Hahn–Banach Theorem; however, such a norming functional is generally not unique. For example, the vector $\mathbf{a} = (1, 1, 0, 0, 0, \ldots) \in c_0$ is normed by both $(1, 0, 0, 0, \ldots)$ and $(0, 1, 0, 0, 0, \ldots)$ in ℓ_1, as well as any convex combination of these two vectors. A norming functional of $\mathbf{x}$, multiplied by the length of $\mathbf{x}$, is called a "conjugate" of $\mathbf{x}$ in some texts, such as [17]. The collection of norming functionals of some nonzero vector $\mathbf{x}$ is sometimes known in the literature as the "peak set" for $\mathbf{x}$; see, for example, [2, p. 939].

The following lemma relates the solutions of the original extremal problem to those of its dual, and thus enables us to drastically narrow our search for the solution set.

Lemma 5.1. *If* $\mathbf{x}_0$ *is a solution to the extremal problem* (2.2), *then* $\mathbf{x}_0/\|\mathbf{x}_0\|_1$ *is a norming functional for any solution* $\mathbf{a}'$ *of the dual problem.*

Proof. Plainly $\mathbf{x}_0/\|\mathbf{x}_0\|_1$ has unit norm. By hypothesis and by Proposition 3.3,

$$\|\mathbf{x}_0\|_1 = \inf\{\|\mathbf{x}_0 + \mathbf{z}\|_1 : \mathbf{z} \in \mathscr{M}\} = \sup_{c_1, c_2, \ldots, c_m} \frac{\sum_{j=1}^m c_j y_j}{\|\sum_{j=1}^m c_j \mathbf{a}_j\|_\infty} \quad \frac{\sum_{j=1}^m c_j' y_j}{\|\sum_{j=1}^m c_j' \mathbf{a}_j\|_\infty}.$$

Hence, we have that

$$\begin{aligned} \langle \mathbf{a}', \mathbf{x}_0/\|\mathbf{x}_0\|_1 \rangle &= \left\langle \sum_{j=1}^m c_j' \mathbf{a}_j, \mathbf{x}_0 \right\rangle \|\mathbf{x}_0\|_1 \\ &= \left(\sum_{j=1}^m c_j' y_j \right) \frac{\|\sum_{j=1}^m c_j' \mathbf{a}_j\|_\infty}{\sum_{j=1}^m c_j' y_j} \\ & \left\| \sum_{j=1}^m c_j' \mathbf{a}_j \right\|_\infty = \|\mathbf{a}'\|_\infty. \end{aligned}$$

According to the definition of the norming functional, we conclude that $\mathbf{x}_0/\|\mathbf{x}_0\|_1$ is a norming functional for $\mathbf{a}'$. □

Notice that because of Lemma 5.1, it is not necessary to find all of the solutions to the dual problem; having one dual solution $\mathbf{a}'$ will suffice for solving the original problem (2.2).

We next describe the norming functionals of a vector in c_0 explicitly. Since the components of $\mathbf{a}$ in c_0 converge to zero, $\|\mathbf{a}\|_\infty = \sup_k |a_k|$ must be attained on a finite set of indices. An index set $\mathscr{N}$ is called the extremal index set for $\mathbf{a} \in c_0$ if $\|\mathbf{a}\|_\infty = \sup_k |a_k|$ is attained on $\mathscr{N}$ and $\mathscr{N}$ is the largest set having this property.

Lemma 5.2. *If a nonzero sequence* $\mathbf{a} \in c_0$ *has its extremal index set given by* $\mathscr{N} := \{n_1, n_2, \ldots, n_N\}$, *then the set of norming functionals for* $\mathbf{a}$ *consists exactly of the convex combinations of vectors of the form*

$$\mathbf{v}_a := \operatorname{sign}(a_j)\mathbf{e}_j, \tag{5.1}$$

where $j \in \mathscr{N}$.

Proof. For each $j \in \mathcal{N}$, the vector $\mathbf{v}_a$ is a unit vector of ℓ_1 such that

$$\langle \mathbf{a}, \mathbf{v}_a \rangle = |a_j| = \|\mathbf{a}\|_\infty;$$

that is, $\mathbf{v}_a$ is a norming functional for $\mathbf{a}$. In fact, any convex combination of such vectors is also norming for $\mathbf{a}$.

Conversely, suppose that $\mathbf{v} \in \ell_1$ is a norming functional for $\mathbf{a}$ and we shall show that $\mathbf{v}$ must be a convex combination of vectors of the form (5.1). By definition, we first observe that $\|\mathbf{v}\|_1 = 1$ and $\langle \mathbf{a}, \mathbf{v} \rangle = \|\mathbf{a}\|_\infty$. Furthermore, we find that

$$\langle \mathbf{a}, \mathbf{v} \rangle = \sum_{j=1}^{\infty} a_j v_j = \sum_{j \in \mathcal{N}} a_j v_j + \sum_{j \notin \mathcal{N}} a_j v_j. \tag{5.2}$$

For the first term of the right-hand side in Eq. (5.2), a direct computation leads to

$$\sum_{j \in \mathcal{N}} (\operatorname{sign}(a_j) a_j)(\operatorname{sign}(a_j) v_j) = \sum_{j \in \mathcal{N}} \|\mathbf{a}\|_\infty \operatorname{sign}(a_j) v_j.$$

That is,

$$\sum_{j \in \mathcal{N}} a_j v_j = \|\mathbf{a}\|_\infty \sum_{j \in \mathcal{N}} \operatorname{sign}(a_j) v_j. \tag{5.3}$$

Our next goal is to show that $v_j = 0$ whenever $j \notin \mathcal{N}$. To accomplish this we note that $\mathbf{a}$, as a real sequence, converges to zero and thus to no other point; in particular, it cannot be that any subsequence converges to $\|\mathbf{a}\|_\infty$. Consequently, we must have

$$\alpha := \sup\{|a_j| : j \notin \mathcal{N}\} < \|\mathbf{a}\|_\infty.$$

Suppose now for the sake of argument that

$$c := \sum_{j \in \mathcal{N}} |v_j| < 1.$$

Notice that $\|\mathbf{v}\|_1 = 1$. It would follow

$$\sum_{j \notin \mathcal{N}} |v_j| = 1 - c \leq 1.$$

This together with (5.3) would imply that

$$\|\mathbf{a}\|_\infty = \sum_{j=1}^{\infty} a_j v_j \leq \left| \sum_{j \in \mathcal{N}} a_j v_j \right| + \left| \sum_{j \notin \mathcal{N}} a_j v_j \right| \leq \|\mathbf{a}\|_\infty c + \alpha(1 - c) < \|\mathbf{a}\|_\infty,$$

an absurdity. This proves that

$$\sum_{j \in \mathcal{N}} |v_j| = 1 = \|\mathbf{v}\|_1, \tag{5.4}$$

and consequently $v_j = 0$ whenever $j \notin \mathcal{N}$. Returning to (5.2) and (5.3), we see that the second summation vanishes, and it must be that

$$\sum_{j \in \mathcal{N}} \operatorname{sign}(a_j) v_j = 1.$$

Combining the equation above with (5.4) yields

$$\sum_{j\in\mathcal{N}} (|v_j| - \mathrm{sign}(a_j)v_j) = 0.$$

This implies that

$$|v_j| = \mathrm{sign}(a_j)v_j, \quad \text{for all } j \in \mathcal{N}.$$

Namely, for all $j \in \mathcal{N}$, the terms $\mathrm{sign}(a_j)v_j$ are nonnegative. From this, it follows that

$$\mathbf{v} = \sum_{j\in\mathcal{N}} v_j \mathbf{e}_j = \sum_{j\in\mathcal{N}} [\mathrm{sign}(a_j)v_j]\mathrm{sign}(a_j)\mathbf{e}_j. \tag{5.5}$$

That is, $\mathbf{v}$ must be a convex combination of vectors of the form (5.1). $\square$

Observe that the collection of norming functionals for a given $\mathbf{a} \in c_0$ constitutes a finite-dimensional "face" or "edge" of a sphere in ℓ_1, as expected. Thus our search for the solution to the original extremal problem is thereby narrowed from the hyperplane $\mathcal{S}$ to those vectors belonging to $\mathcal{S}$ that are supported on the finite extremal index set $\mathcal{N}$, and which take the form (5.5). This is a significant reduction in the scope of the search. It remains to find the coefficients in this representation such that $\mathbf{v}$ is a solution.

Remark 5.3. The formula (5.5) could be viewed as a kind of Representer Theorem, in which the solution to the ℓ_1 interpolation problem is expressed as certain a finite-dimensional linear combination.

Among the vectors satisfying (5.5) there are some that also satisfy the system (2.3). To find them, associated with the extremal index set $\mathcal{N} := \{n_1, n_2, \ldots, n_N\}$ for a dual solution $\mathbf{a}' \in c_0$ as described in Lemma 5.2, we first define an infinite permutation matrix $\mathbf{Q}$ (it is the infinite identity matrix with finitely many of the columns permutated) that, when acting on a column vector, interchanges the n_kth row with the kth row, for $k = 1, 2, \ldots, N$, and affects N such pairs of rows.

Using the permutation matrix $\mathbf{Q}$, the original system (2.4) can then be reexpressed in matrix form as

$$(\mathbf{AQ})(\mathbf{Qx}) = \mathbf{y}$$

where A is again the $m \times \infty$ matrix with rows being the $\mathbf{a}_j$ sequences; $\mathbf{x}$ is a ∞-dimensional column vector; and $\mathbf{y}$ is an m-dimensional column vector.

Let $\hat{\mathfrak{x}}$ be the N-dimensional column vector consisting of the first N entries of $\mathbf{Qx}$ (by choice its remaining entries are all zeros), and let $\mathbf{B}$ be the $m \times N$ rectangular matrix consisting of the N leftmost columns of $\mathbf{AQ}$. It is elementary to solve the system

$$\mathbf{B}\hat{\mathfrak{x}} = \mathbf{y} \tag{5.6}$$

for the N-dimensional vector $\hat{\mathfrak{x}}$. This is a *finite* matrix algebra problem, and numerous techniques exist for computing the solution. A solution for $\hat{\mathfrak{x}}$ exists; this is because if infinitely many zeros are appended to $\hat{\mathfrak{x}}$ to make it an infinite column vector, then some $\mathbf{Q}\hat{\mathfrak{x}}$ must be a solution to (2.3). Let $\mathscr{H}$ be the solution set for the finite-dimensional system (5.6) and let

$$\mathscr{I} := \mathscr{H} \cap \left\{ (w_1, w_2, \ldots, w_N) \in \mathbb{R}^N : \sum_{j=1}^{N} |w_j| = m_S \right\}.$$

In fact for $(w_1, w_2, \ldots, w_N) \in \mathscr{I}$, we must have that

$$\sum_{j=1}^{N} |w_j| = m_S = \sum_{j=1}^{N} \operatorname{sign}(a_{n_j}) w_j,$$

in relation to the notation of (5.5). We are in effect narrowing the search further from a finite-dimensional subset of the hyperplane $\mathscr{S}$ to its intersection with the sphere in ℓ_1 with radius m_S. This intersection is necessarily confined to a "face" or "edge" of this sphere, reflecting the form of (5.7). Every vector belonging to the set $\mathbf{Q}\mathscr{I}$ is a solution $\mathbf{x}_0$ to the infimum problem (2.2). Each such solution vector is supported on a finite collection of indices; that is, it is a sparse vector.

The above discussion establishes the following theorem.

Theorem 5.4. *Let $\mathbf{a}' \in c_0$ be a solution of the dual problem* (3.5), $\mathbf{Q}$ *the permutation matrix associated with the extremal index set for* $\mathbf{a}'$, *and* $\mathbf{B}$ *the* $m \times N$ *matrix, as defined above. Then solutions to the minimum norm ℓ_1 interpolation problem* (2.2) *consist of those vectors* $\mathbf{Qx}$, *where* $\|\mathbf{Qx}\|_1 = m_S$, *the vector consisting of the first N components of* $\mathbf{x}$ *solves the finite-dimensional matrix equation* (5.6), *and the remaining entries of* $\mathbf{x}$ *are zero.*

Let us summarize the final stage of this solution method as follows.

Step 1: For a given solution $\mathbf{a}' \in c_0$ of the dual problem (3.5), construct the extremal index set $\mathscr{N} := \{n_1, n_2, \ldots, n_N\}$ for $\mathbf{a}'$, necessarily a finite set.

Step 2: Based on the extremal index set $\mathscr{N}$, define the infinite permutation matrix $\mathbf{Q}$ that interchanges the kth row with the n_kth row, for every $n_k \in \mathscr{N}$.

Step 3: Solve the finite-dimensional linear algebra equation (5.6), thus obtaining a set $\mathscr{H}$ of N-dimensional vectors.

Step 4: Among the vectors in $\mathscr{H}$, identify those of length m_S, the extremal value previously obtained in Sec. 4. This effort involves solving a single equation in at most N variables, and results in a bounded, convex subset $\mathscr{I}$ of $\mathbb{R}^N$.

Step 5: Re-embedding of the members of $\mathscr{I}$ back into ℓ_1 using $\mathbf{Q}$ yields the complete set of solution vectors to the minimum norm interpolation problem (2.2).

We have thus shown that the original interpolation problem in ℓ_1 is equivalent to solving a linear programming problem derived from a duality argument, followed by solving a finite-dimensional linear matrix equation, resulting in a sparse solution.

Remark 5.5. This paper sets forth a conceptual road map for solving the ℓ_1 minimum norm interpolation problem (2.2). In practice, consideration must be given to the computational complexity and stability of the solution method. For example, computational complexity will increase with the parameter m, the number of given vectors in the subset S of c_0; the number of vertices of the polytope Δ arising in the associated linear programming problem; the parameter N, the number of dimensions of the edge or face of the sphere in ℓ_1 in the final reduction. Concern for the stability of the solution arises in connection with solving the dual extremal problem, identifying the extremal index set, and solving the finite-dimensional linear algebra equation, and imposing the minimum length condition on the resulting vectors. These issues relating to the implementation of the solution method will be addressed in forthcoming research.

6. Example

We now illustrate the method developed in this paper by solving a simple but nontrivial example.

The number of constraints in the initial interpolation problem will be $m = 2$. Fix

$$y_1 = 3, \quad y_2 = 4, \tag{6.1}$$

$$\mathbf{a}_1 = \left(1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots\right) = \left(\frac{1}{n}\right)_{n=1}^{\infty}, \tag{6.2}$$

$$\mathbf{a}_2 = \left(1, -\frac{1}{2}, \frac{1}{4}, -\frac{1}{8}, \dots\right) = \left(\frac{1}{[-2]^{n-1}}\right)_{n=1}^{\infty}. \tag{6.3}$$

We consider the problem of finding $\mathbf{x}_0 \in \ell_1$ such that

$$\|\mathbf{x}_0\|_1 = \inf\{\|\mathbf{x}\|_1 : \langle \mathbf{a}_i, \mathbf{x}\rangle = y_i, i = 1, 2\}.$$

This is the main interpolation problem in ℓ_1 from (2.2).

The corresponding dual extremal problem (3.5) is to find $\mathbf{c} \in \mathbb{R}^2$ which attains

$$\sup_{c_1, c_2} \frac{c_1 y_1 + c_2 y_2}{\|c_1 \mathbf{a}_1 + c_2 \mathbf{a}_2\|_\infty} = \sup_{c_1, c_2} \frac{c_1 y_1 + c_2 y_2}{\|(c_1, c_2)\|_\star}.$$

To solve the dual extremal problem we must look at the closed unit ball in $\mathbb{R}^2$ in the $\|\cdot\|_\star$ norm. This unit ball consists of the intersection of infinite strips

$$U_k := \{\mathbf{c} \in \mathbb{R}^2 : -1 \le c_1 a_{1,k} + c_2 a_{2,k} \le 1\} = \left\{\mathbf{c} \in \mathbb{R}^2 : -1 \le \frac{c_1}{k} + \frac{c_2}{(-2)^{k-1}} \le 1\right\}$$

over all $k = 1, 2, \ldots$. To find $U_1 \cap U_2$, we solve the system

$$c_1 \cdot 1 + c_2 \cdot 1 = \pm 1,$$

$$c_1 \cdot \frac{1}{2} + c_2 \cdot \left(-\frac{1}{2}\right) = \pm 1$$

for all choices of sign. We thereby obtain the convex polytope (in this case, a polygon) with vertices at

$$\left(-\frac{1}{2}, \frac{3}{2}\right), \left(\frac{3}{2}, -\frac{1}{2}\right), \left(\frac{1}{2}, -\frac{3}{2}\right), \left(-\frac{3}{2}, \frac{1}{2}\right).$$

In this example, the intersection over all the U_k turns out to be equal to $U_1 \cap U_2 = \Delta$. To see this, first note that each vertex of $U_1 \cap U_2$ lies at a distance

$$\sqrt{\left(\frac{1}{2}\right)^2 + \left(\frac{3}{2}\right)^2} = \left(\frac{1}{2}\right)\sqrt{10} \approx 1.581$$

from the origin. On the other band, the strip U_3 lies at a distance

$$\frac{1}{\sqrt{\left(\frac{1}{3}\right)^2 + \left(\frac{1}{4}\right)^2}} = 2.4 > 1.581$$

from $\mathbf{0}$. That is, U_3 encloses $U_1 \cap U_2$ completely, and so $U_3 \cap (U_1 \cap U_2) = U_1 \cap U_2$. The strips U_j, for $j > 3$, are even wider still, and therefore do not contribute to defining Δ further. This verfies that $U_1 \cap U_2 = \Delta$. Thus to solve the dual extremal problem, we are finding the maximum of a linear function of c_1 and c_2, confined to the boundary of the rectangle Δ. We know from Theorem 4.4 that it suffices to plug in the vertices (c_1, c_2) of Δ into the function

$$f(x_1, x_2) := 3x_1 + 4x_2, \quad (x_1, x_2) \in \Delta,$$

and compare. The maximum value occurs at the vertex $(-\frac{1}{2}, \frac{3}{2})$, with the maximum value being

$$m_S = f\left(-\frac{1}{2}, \frac{3}{2}\right) = \frac{9}{2}.$$

Our next step is to identify the (necessarily finitely many) indices for which the sequence

$$\mathbf{a}' = -\frac{1}{2}\mathbf{a}_1 + \frac{3}{2}\mathbf{a}_2$$

attains its supremum norm. By direct computation we find that

$$\left|-\frac{1}{2}a_{1,1} + \frac{3}{2}a_{2,1}\right| = \left|-\frac{1}{2} + \frac{3}{2}\right| = 1,$$

$$\left|-\frac{1}{2}a_{1,2} + \frac{3}{2}a_{2,2}\right| = \left|-\frac{1}{4} - \frac{3}{4}\right| = 1,$$

$$\left|-\frac{1}{2}a_{1,k} + \frac{3}{2}a_{2,k}\right| = \left|-\frac{1}{2k} + \frac{3}{(-2)^k}\right| < 1, \quad \text{for all } k > 2.$$

This tells us that the extremal index set $\mathscr{N}$ we are seeking is $\{1,2\}$. The solution of the original extremal problem must be supported on these two indices. Since they already correspond to the two leftmost components of a vector in c_0, the permutation matrix $\mathbf{Q}$ occurring in this example is simply the identity. We also infer from the above computation that

$$\mathbf{a}' = \left(1, -1, -\frac{13}{24}, \dots\right).$$

We may therefore truncate the system $\mathbf{A}\mathbf{x} = \mathbf{y}$ to get

$$\begin{bmatrix} 1 & \frac{1}{2} \\ 1 & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}, \tag{6.4}$$

corresponding to Eq. (5.6). We have thereby reduced the original infinite-dimensional interpolation problem to a routine linear algebra problem.

The system (6.4) has the solution set $\mathscr{H} = \{[\frac{7}{2}, -1]^T\}$. Since the solution set is a single vector $\hat{\mathbf{x}} = [\frac{7}{2}, -1]^T$ in this example, our search is complete. We re-embed this vector $\hat{\mathbf{x}}$ into ℓ_1, and obtain the following solution to the original extremal problem:

$$\mathbf{x}_0 = \left(\frac{7}{2}, -1, 0, 0, 0, \dots\right).$$

Notice that for the extreme value we get $\|\mathbf{x}_0\|_1 = m_S = \frac{9}{2}$, in agreement with the dual problem as expected. Furthermore, we can confirm by inspection that $\mathbf{x}_0/\|\mathbf{x}_0\|_1$ is norming for $\mathbf{a}'$, also as expected.

For the sake of comparison, here is the solution to the same example, except we are using the norm of ℓ_2. With y_1, y_2, $\mathbf{a}_1$ and $\mathbf{a}_2$ defined as in (6.1)–(6.3), we are seeking the $\mathbf{x}_0 \in \ell_2$ for which

$$\|\mathbf{x}_0\|_2 = \inf\{\|\mathbf{x}\|_2 : \langle \mathbf{a}_i, \mathbf{x}\rangle = y_i, i = 1, 2\}.$$

Here, the notation $\langle \cdot, \cdot \rangle$ denotes the usual inner product in the Hilbert space ℓ_2. If $\mathbf{x}'$ is any particular vector satisfying $\langle \mathbf{a}_i, \mathbf{x}'\rangle = y_i, i = 1, 2$, then equivalently we are seeking to minimize $\|\mathbf{x}' + \mathbf{z}\|_2$ over all $\mathbf{z}$ lying in the subspace of ℓ_2 annihilated by $\mathbf{a}_1$ and $\mathbf{a}_2$. This exactly describes the orthogonal projection of $\mathbf{x}'$ onto the span of $\mathbf{a}_1$ and $\mathbf{a}_2$ in ℓ_2. Thus if $\mathbf{u}_1$ and $\mathbf{u}_2$ constitute an orthonormal basis for the subspace spanned by $\mathbf{a}_1$ and $\mathbf{a}_2$, then we have

$$\mathbf{x}_0 = \langle \mathbf{x}', \mathbf{u}_1\rangle \mathbf{u}_1 + \langle \mathbf{x}', \mathbf{u}_2\rangle \mathbf{u}_2.$$

The choice

$$\mathbf{x}' = \left(\frac{7}{2}, -1, 0, 0, 0, \dots\right)$$

will suffice, and yields the ℓ^2 solution

$$x_0 \approx (0.4924584)\left(\frac{1}{n}\right)_{n=1}^{\infty} + (2.7004714)\left(\frac{1}{(-2)^{n-1}}\right)_{n=1}^{\infty}$$
$$\approx (3.1929568, -1.1039930, 0.8392707, -0.2144443, \ldots).$$

This solution is certainly not a sparse vector.

This example shows that the minimum norm interpolation problem in the $\ell_1(\mathbb{N})$ space indeed produces a sparse solution, while that in the $\ell_2(\mathbb{N})$ space does not.

Acknowledgments

This work is supported in part by the US National Science Foundation under Grant DMS-1912958. Y. Xu is also a professor emeritus of mathematics at Syracuse University, New York, USA, and a retired professor at Sun Yat-sen University, Guangzhou, China. All correspondence should be sent to Y. Xu.

References

[1] B. Adcock, Infinite-dimensional compressed sensing and function interpolation, *Found. Comput. Math.* 18 (2018) 661–701.
[2] A. Argyriou, C. A. Micchelli and M. Pontil, On spectral learning, *J. Mach. Learn. Res.* 11 (2010) 935-953.
[3] D. Avis and K. Fukuda, A pivoting algorithm for convex hulls and vertex enumeration of arrangements and polyhedra, *Discrete Comput. Geom.* 8 (1992) 295-313.
[4] E. J. Candès, J. Romberg and T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, *IEEE Trans. Inform. Theory* 52 (2006) 489–509.
[5] N. L. Carothers, *A Short Course on Banach Space Theory* (Cambridge University Press, Cambridge, 2005).
[6] J. Conway, *A Course in Functional Analysis* (Springer-Verlag, New York, 1985).
[7] F. Cucker and S. Smale, On the mathematical foundations of learning, *Bull. Amer Math. Soc.* 39 (2002) 1-49.
[8] F. Deutsch, V. A. Ubhaya, J. D. Ward and Y. Xu, Constrained best approximation in Hilbert space III: Applications to n-convex functions, *Constr. Approx.* 12 (1996) 361–384.
[9] D. L. Donoho, Compressed sensing, *IEEE Trans. Inform. Theory* 52 (2006) 1289–1306.
[10] T. Evgeniou, M. Pontil and T. Poggio, Regularization networks and support vector machines, *Adv. Comput. Math.* 13 (2000) 1–50.
[11] B. Grünbaum, *Convex Polytopes*, 2nd edn. (Springer-Verlag, New York, 2003).
[12] Katta G. Murty, *Linear Programming* (John Wiley & Sons, New York, 1983).
[13] G. Nielson, Minimum norm interpolation in triangles, *SIAM J. Numer. Anal.* 17 (1980) 44-62.
[14] M. Reed and B. Simon, *Functional Analysis* (Academic Press, New York, 1980).
[15] G. Song and H. Zhang, Reproducing kernel Banach spaces with the ℓ_1 norm II: Error analysis for regularized least square regression, *Neural Comput.* 23 (2011) 2713–2729.
[16] G. Song, H. Zhang and F. J. Hickernell, Reproducing kernel Banach spaces with the ℓ_1 norm, *Appl. Comput. Harmon. Anal.* 34 (2013) 96-116.

[17] M. Unser, A unifying representer theorem for inverse problems and machine learning, preprint.
[18] M. Unser, J. Fageot and H. Gupta, Representer theorems for sparsity-promoting ℓ_1 regularization, *IEEE Trans. Inform. Theory* 63 (2016) 5167–5180.
[19] Robert J. Vanderbei, *Linear Programming: Foundations and Extensions* (Springer-Verlag, New York, 2001).
[20] R. Wang and Y. Xu, Functional reproducing kernel Hilbert spaces for non-point-evaluation functional data, *Appl. Comput. Harmon. Anal.* 46 (2019) 569–623.
[21] Y. Xu and Q. Ye, Generalized Mercer kernels and reproducing kernel Banach spaces, *Mem. Amer. Math. Soc.* 258 (2019) 1–122.
[22] H. Zhang, Y. Xu and J. Zhang, Reproducing kernel Banach spaces for machine learning, *J. Mach. Learn. Res.* 10 (2009) 2741–2775.