Deconvolution Filter Design via $\ell_1$ Optimization Technique

Sen-Chueh Peng, Member, IEEE and Bor-Sen Chen, Senior Member, IEEE

Abstract—A new $\ell_1$ optimal deconvolution filter design approach for systems with uncertain (or unknown)-but-bounded inputs and external noises is proposed. The purpose of this deconvolution filter is to minimize the peak gain from the input signal and noise to the error by the viewpoint of the time domain. The solution consists of two steps. In the first step, the $\ell_1$ norm minimization problem is transferred to an equivalent $A$-norm minimization problem, and the minimum value of the peak gain is calculated. In the second step, based on the minimum peak gain, the $\ell_1$ optimal deconvolution filter is constructed by solving a set of constrained linear equations. Some techniques of inner-outer factorization, polynomial spectral factorization, linear programming, and some optimization theorems found in a book by Luenberger are applied to treat the $\ell_1$ optimal deconvolution filter design problem. Although the analysis of the algorithm seems complicated, the calculation of the proposed design algorithm for actual systems is simple. Finally, one numerical example is given to illustrate the proposed design approach. Several simulation results have confirmed that the proposed $\ell_1$ optimal deconvolution filter has more robustness than the $\ell_2$ optimal deconvolution filter under uncertain driving signals and noises.

I. INTRODUCTION

The deconvolution problem is widely found in engineering articles [1]–[5]. The problem differs from the usual filtering problems because it not only estimates a signal embedded in noise, but it also removes the effect of any distortion in the channel systems. The applications of the deconvolution filter are to detect signals corrupted with channel noise and remove the intersymbol effect of the channel in communication systems [1], [2], as well as to estimate a signal from the seismic trace contaminated by noise and remove the effect of distortion of the seismic wavelet [3]–[5]. However, the deconvolution problem is complicated in at least two respects: The measurements are usually corrupted by noise, and the system is frequently nonminimum phase. These problems restrict the use of the simplest deconvolution filter, namely, the inverse filter.

During the past two decades, there have been several methods proposed to treat this deconvolution problem. An optimal deconvolution filter was first designed via least square predictions error by Silvia and Robinson [4]. However, only the anticausal deconvolution filter is obtained in a single channel system. Mendel [6] has proposed a state space solution to the fixed-lag minimum variance deconvolution (MVD) problem. The solution involves passing the data through a Kalman filter followed by processing (off line) the innovation sequence with an anticausal filter. A detailed comparison between various deconvolution techniques has been proposed by Chi and Mendel [7]. They have shown that the steady-state Kalman filter is equivalent to the prediction error filter of Silvia and Robinson [4]. In addition, regarding the deconvolution problem, Moir [8] has provided a solution that can be found in transfer function matrix form and in polynomial matrix form. Ahlén and Sternad [9], as well as Chisci and Mosca [10], have solved the optimal deconvolution filter problem based on the spectral factorization of polynomial and diophantine equations. Chen and Peng [11], based on the orthogonal principle, have also provided a solution to the deconvolution problem.

The design objective of those methods is to minimize a quadratic integral-type criterion, either in terms of the frequency domain such as Wiener optimization or in the light of the time domain such as Kalman filtering. The main advantages of using quadratic-type criteria is that they contain the explicit solution to the corresponding optimal deconvolution problem and the accessibility for analysis. Although the optimization with quadratic-type criteria is a powerful tool, it has limitations. When the environment has some uncertainties in terms of driving input or external noise, which usually occur, the performance of the optimal deconvolution filter associated with some nominal system will be degraded. To overcome the limitations, adaptive deconvolution schemes can be applied to deal with the deconvolution problem. In the past, several adaptive deconvolution algorithms have been proposed [12]–[14]. Although adaptive deconvolution schemes are occasionally applied, they may be impractical because of cost and complexity.

Thus, in recent years, the need for robust estimators has led to the development of a large number of useful schemes. The minimax approaches for robust filtering design have been given by [15] and [16]. The minimax approaches to the design of filters that are robust with respect to noises and uncertainties of parameters are based on a game theory. In addition, the $H_\infty$ optimization approach proposed by [17] is used to treat the deconvolution problem in the frequency domain for giving an acceptable performance over an entire range of uncertainties resulting from driving inputs and external noises. Whether the criterion is of the quadratic type or in $H_\infty$-optimization, the design philosophy of the algorithms of [4], [6], [7], [9]–[11], and [17] is that the system inputs and external noises are considered to have finite energy, i.e., bounded in $\ell_2$ norm. In
practice, inputs and noises usually do not satisfy this condition and act more or less continuously over time. In this case, such inputs and noises are said to be persistent and cannot be treated in the $\ell_2$ optimization framework. Hence, the approaches of [4], [6], [7], [9]–[11], and [17] clearly do not apply if the infinite energy signals are to be estimated.

One of the most recent and interesting developments in control theory is the application of $\ell_1$ controllers for uncertain-but-bounded inputs or external noises from the viewpoint of minimizing the peak gain of an output error in the time domain [18]. Although the design of the $\ell_1$ deconvolution filter can be regarded as a dual problem for the design of $\ell_1$ controllers, it has received little attention. Additionally, the $\ell_1$ deconvolution filter in this paper is different from the conventional $\ell_1$ deconvolution filter given by [19] and [20]. The main differences between them are twofold. The conventional $\ell_1$ deconvolution filter merely deals with FIR noiseless cases, and it may lack stability, but the proposed $\ell_1$ deconvolution filter proposed here is able to deal with IIR noise-corrupted cases, and it will be ascertained to be stable. During the reviewing process of this paper, Mendlovitz [21] presented an $\ell_1$ optimal estimator by solving a sequence of suboptimization and superoptimal linear programming methods. The weakness of this method is that the number of constraints on suboptimal and superoptimal linear programming may be infinite because the $\ell_1$ optimal estimator is of FIR form and must be of infinite order. Although some order truncation has been used, however, the more the accuracy of the estimator is required, the higher order of the estimator is necessitated. Consequently, the estimator may still be of relatively high order.

The purpose of this paper is to study the single-channel deconvolution problem by utilizing the $\ell_1$ optimization theory in the time domain. A corresponding $\ell_1$ optimal deconvolution filtering design algorithm is then presented to realize a causal and stable deconvolution filter for both minimum- and nonminimum-phase channel systems. However, the problem of deconvolution is to investigate how to use only one filter to remove the distortion in the channel (interference intersymbol) and simultaneously reject the corrupted noise. The optimal deconvolution problem will lead to a two-block (nonsquare) $\ell_1$ optimization problem, which is more difficult to solve than the one-block (square) $\ell_1$ optimization problem in discrete feedback control design [18]. At first, the $\ell_1$ optimization problem is transformed to an equivalent A-norm optimization problem. Inner–outer factorization technique [23], [24], is used to solve the so-called interpolation problem of nonminimum zeros and to make the order of optimal deconvolution filter as small as possible. To solve the deconvolution filtering problem, some techniques, such as the linear programming technique [22], polynomial spectral factorization [25], and some optimization theorems found in [26] are utilized. Furthermore, in our method, a closed-form IIR solution can be obtained by one iteration of linear programming. To illustrate, a numerical design example for a discrete-time single-channel system contaminated by noise is considered.

The remainder of this paper is organized as follows. In Section II, the mathematical preliminaries are described. In Section III, the deconvolution problem contaminated by noise is formulated. The solution to the deconvolution problem for the nonminimum-phase system and the corresponding design algorithm are presented in Section IV. A simulation result is presented in Section V to illustrate the proposed design procedure. Finally, conclusions are given in Section VI.

Before ending the introduction, we shall define some of the notations being used in this paper:

\[ \begin{align*}
X & \quad \text{Real normed linear space.} \\
X^* & \quad \text{Dual space of } X. \\
B^\perp & \quad \text{Orthogonal complement of } B (B^\perp \subset X^*). \\
\ell_1^p & \quad \text{Space of all real sequences with } \| \cdot \|_p < \infty. \\
\ell_p^N & \quad \text{Space of } N\text{-length sequences with the } \ell_p\text{-norm.} \\
\ell_0 & \quad \text{Subspace of } \ell_\infty \text{ with } \{ h \in \ell_\infty : \sum_{k=0}^{\infty} |h_k| < \infty \}. \\
\ell_0^c & \quad \text{Subspace of } \ell_\infty \text{ with } \{ h \in \ell_\infty : \lim_{k \to \infty} h_k = 0 \}. \\
A & \quad \text{Space of } Z\text{-transforms of all sequences in } \ell_1 \text{ Norm defined on the space } A. 
\end{align*} \]

II. PRELIMINARIES

The purpose of this section is to fix the notations and provide some definitions. Given a sequence \( f = \{f(k)\} \), we can define two norms on it as follows:

\[ \begin{align*}
\|f\|_1 &= \sum_{k=0}^{\infty} |f(k)|, \\
\|f\|_\infty &= \sup_{k} |f(k)|. 
\end{align*} \] (2.1)

The sequence spaces \( \ell_1, \ell_\infty \) are defined, respectively, to consist of those sequences \( \{f(k)\} \) such that \( \|f\|_1, \|f\|_\infty \) is finite. Then, the \( Z \)-transform of the given sequence \( f \) is defined as

\[ F(z^{-1}) = \sum_{k=0}^{\infty} f(k)z^{-k} \] (2.3)

where \( z^{-1} \) is a backward-shift operator. The current definition has some advantages, one of which is that every finite-length polynomial in the indeterminate \( z \) with real coefficients belongs to \( A \). In fact, such a finite-length polynomial corresponds to a so-called finite impulse response (FIR) filter.

Suppose \( h \in \ell_1 \) associates with the impulse response of a linear operator denoted by \( H \), which maps a sequence \( f \in \ell_\infty \) into its convolution with \( h \). That is

\[ H := \ell_\infty \rightarrow \ell_\infty, \quad Hf = h * f. \nonumber \]

The induced norm of \( H \) is given by [18], [26]

\[ \|H\| := \sup_{f \in \ell_\infty} \frac{\|h * f\|_\infty}{\|f\|_\infty} = \sup_{\|f\|_\infty = 1} \|h * f\|_\infty = \|h\|_1. \] (2.4)

It is evident that the induced norm of \( H \), which corresponds to the peak gain of operator \( H \), is equal to the \( \ell_1 \) norm of its impulse response. The \( Z \)-transform of \( h \) is denoted by \( H(z^{-1}) \in A \), and the norm on \( A \) is defined as follows [18]:

\[ \|H(z^{-1})\|_A := \|h\|_1. \] (2.5)
Let \( b_0 \) denote the subspace of \( \ell_\infty \) consisting of all sequences that are absolutely summable. It is seen that \( b_0 \) and \( \ell_1 \) are distinct normed spaces whose underlying linear vector spaces are the same. In other words, \( b_0 \) and \( \ell_1 \) are algebraically identical but topologically different. Consequently, the \( Z \) transforms of all sequences in \( b_0 \) are also stable. Note that hereafter, all elements of \( A \) and all \( Z \) transforms of \( b_0 \) sequences are rational unless explicitly stated to the contrary.

Next, some definitions are given as follows. Some optimization theorems found in [26] are utilized extensively in the analysis of our problems and will be stated in the Appendix.

**Definition 1:** Let \( x \in X \) and \( x^* \in X^* \); then, the notation \( \langle x, x^* \rangle \) is for the value of the functional \( x^* \) at a point \( x \in X \). The norm of an element \( x^* \in X^* \) is

\[
||x^*||_B = \sup_{x \neq 0} \frac{|\langle x, x^* \rangle|}{||x||_B}
\]

where \( ||\cdot||_B \) (resp., \( ||\cdot||_Q \)) denotes the norm on \( X \) (resp., \( X^* \)).

**Definition 2:** Let \( B \) be a subspace of \( X \). The orthogonal complement of \( B \), which is denoted by \( B^\perp \), is defined as follows:

\[
B^\perp = \{ x^* \in X^* : \langle x, x^* \rangle = 0 \text{ for all } x \in B \}.
\]

**Definition 3:** A vector \( x^* \in X^* \) is said to be aligned with a vector \( x \in X \) if \( \langle x, x^* \rangle = ||x||_B||x^*||_Q \).

**Definition 4:**

i) A stable rational function \( B(z^{-1}) \in A \) is called an inner function if and only if \( B^\top(z)B(z^{-1}) = I \) for almost all \( |z| = 1 \), where the superscript \( \top \) denotes the transpose.

ii) A stable rational function \( C(z^{-1}) \in A \) is called an outer function if and only if all poles and zeros of \( C(z^-1) \) are in \( |z| \leq 1 \).

In the sequel, we will need the concepts of inner and outer matrices. Below, we summarize some definitions from [23]–[25].

**Definition 5:**

i) A stable rational matrix \( G_{n \times m}(z^{-1}) \), \( n \geq m \) is inner if \( G(z)G(z^{-1}) = I_m \) for almost all \( |z| = 1 \). It is co-inner if \( n \leq m \) and \( G(z^{-1})G(z) = I_n \) for almost all \( |z| = 1 \).

ii) A stable rational matrix \( G_{n \times m}(z^{-1}) \), \( n \leq m \) is outer if and only if it has full row rank \( n \forall |z| \geq 1 \). In other words, it has no zeros in \( |z| \geq 1 \). It is co-outer when \( n \geq m \) if and only if it has full column rank \( m \forall |z| \geq 1 \).

iii) A stable rational matrix \( G_{n \times m}(z^{-1}) \), with full rank \( p \equiv \max \{ m, n \} \) for all \( z = e^{\omega t} \) (no zeros on the unit circle) has an inner–outer factorization

\[
G_{n \times m}(z^{-1}) = G_{n \times p}(z^{-1})G_{p \times m}(z^{-1}) \quad (2.6)
\]

with the outer factor \( G_{p \times m}(z^{-1}) \) having a stable right inverse. It also has a co-inner–outer factorization

\[
G_{n \times m}(z^{-1}) = G_{n \times p}(z^{-1})G_{p \times m}(z^{-1}) \quad (2.7)
\]

with the co-outer factor \( G_{p \times m}(z^{-1}) \) having a stable left inverse. If \( n \leq m \), the co-outer matrix is square, and its inverse is unique.
where $T(q^{-1}) := [L(q^{-1}) - F(q^{-1})H(q^{-1})L(q^{-1}) - F(q^{-1})D(q^{-1})]$ denotes the transfer function matrix from $d(k)$ to $e(k)$ and is nonsquare.

If the considered channel is minimum-phase and noise-free, the desired estimate $\hat{u}(k)$ is the replica of the input signal $u(k)$. Thus, the deconvolution filter $F(q^{-1})$ is easily constructed by the inverse of channel function. If the channel is neither minimum-phase nor noise-free, the deconvolution filter is unable to be directly constructed by the inverse of channel function. The main reasons are that $T(q^{-1})$ is nonsquare, and the inverse of $H(q^{-1})$ is unstable. In this study, since the driving signal $d(k) = [v(k) \ v(k)^T]$ is assumed to be uncertain but bounded, a reasonable design objective is to determine a causal and stable deconvolution filter $F(q^{-1})$ to make the peak gain, from the driving signal $d(k)$ to the estimated error $e(k)$, as small as possible. More detailed analysis is given as follows.

Since $d(k)$ is assumed uncertain but bounded, then from (2.4), (2.5), and (3.4), it is reasonable to take the performance index as [26], [27]

$$
\sigma = \sup_{d \in \mathbb{C}^N, \|d\|_\infty, \text{det}(d) \neq 0} \|t * d\|_\infty = \sup_{\|d\|_\infty} \|t * d\|_\infty
$$

(3.5)

where $t$ denotes the impulse response of $T(q^{-1})$ in (3.4), and its Z transform is $T(z^{-1})$. It is evident that $\sigma$ is the peak gain for the system in (3.4) for all possible uncertain-but-bounded input $d$, and it is equivalent to the $\ell_1$ norm of the impulse response of the system in (3.4). It is a worst-case scenario from input/error peak gain in time domain. Hence, this design criterion is more suitable (or robust) for the deconvolution system with uncertain (or a variety of) inputs and noises. Using (2.5) and (3.4), we further have

$$
\sigma = \|T(z^{-1})\|_A = \|[L(z^{-1}) - F(z^{-1})H(z^{-1})L(z^{-1}) - F(z^{-1})D(z^{-1})]\|_A.
$$

(3.6)

Note that in the time domain, the peak gain $\sigma$ of the system in (3.4) is finite if and only if the inverse transfer function $T(z^{-1})$ is stable.

**Remark 1:** The peak gain of a system can also be expressed in terms of its step response $s(k)$, i.e.

$$
\sigma = TV[s(k)]
$$

where $TV(f)$, which is the total variation of a function $f$, is defined by [26], [27]

$$
TV(f) := \sup_{k=1}^\infty |f(k) - f(k+1)|.
$$

Generally speaking, $TV(f)$ is the sum of all of the consecutive peak-to-valley differences in $f$.

If the considered problem is noise-free or noiseless, then the performance index can be reduced to the following form:

$$
\sigma = \|L(z^{-1}) - F(z^{-1})H(z^{-1})L(z^{-1})\|_A.
$$

(3.7)

The next section will provide a method to construct a causal and stable deconvolution filter $F(z^{-1})$ to minimize $\sigma$ when the inputs and the external noises are uncertain but bounded.

## IV. Optimal Deconvolution Filtering Design

The optimal deconvolution filter design problem of minimizing (3.6) is to minimize the peak gain for the system from $d(k)$ to $e(k)$ when the system’s inputs and external noises are uncertain but bounded in amplitude. Let $T_1(z^{-1}) = [L(z^{-1}) \ 0] \in A$ and $T_2(z^{-1}) = [H(z^{-1})L(z^{-1}) \ D(z^{-1})] \in A$. Both of them are given stable row vectors. Thus, the optimal problem in (3.6) can be expressed in the form

$$
\sigma_0 = \inf_{F(z^{-1}) \in A} \|T_1(z^{-1}) - F(z^{-1})T_2(z^{-1})\|_A
$$

where $t_1, t_2$ are the row vector sequences whose elements are the inverse Z transforms of the elements of $T_1(z^{-1})$, $T_2(z^{-1})$, respectively, and $f$ is the inverse Z transform of $F(z^{-1})$. The purpose of this section is to obtain a stable and causal deconvolution filter $F(z^{-1})$ so that $\|T_1(z^{-1}) - F(z^{-1})T_2(z^{-1})\|_A$ achieves its minimum $\sigma_0$.

When a co-inner–outer factorization is performed based on [25] so that $T_2(z^{-1}) = T_{2CO}(z^{-1}) T_{2CF}(z^{-1})$, where

$$
T_{2CO}(z^{-1}) = \frac{L_n(z^{-1})H_n(z^{-1})D_n(z^{-1})}{\rho(z^{-1})},
$$

and the stable spectral factor $\rho(z^{-1})$ is obtained from the following polynomial spectral factorization [25]:

$$
\rho(z^{-1}) = \frac{L_n(z^{-1})H_n(z^{-1})D_n(z^{-1})}{\rho(z^{-1})}
$$

and

$$
\sigma_0 = \inf_{F(z^{-1}) \in A} \|T_1(z^{-1}) - U(z^{-1})T_{2CF}(z^{-1})\|_A
$$

(4.2)

where $U(z^{-1}) = F(z^{-1})T_{2CO}(z^{-1})$, and $T_{2CF}(z^{-1}) \in A$. Suppose $T_{2CF}(z^{-1})$ has $n_3$ nonminimum phase zeros, $a_i; \ i = 1, \ldots, n_3$, and no zeros on the unit circle. For simplicity, we will assume that $a_i$’s are distinct real. The case of $a_i$ with a complex number or multiplicity will be discussed later.

Let $G(z^{-1}) = F(z^{-1})T_{2CO}(z^{-1}) = U(z^{-1})T_{2CF}(z^{-1}) \in A$. The next lemma determines the conditions on a stable row vector $G(z^{-1})$ such that $U(z^{-1}) \notin T_{2CF}(z^{-1})$ will always be stable; the proof of this lemma is referred to in [18].
Lemma 1 [18]:
1) Given $T_{2C}(z^{-1})$ as the above, there exists a nonzero vector $\beta_k$ such that
$$T_{2C}(a_k)\beta_k = 0 \quad \text{for} \ i = 1, \ldots, n_1.$$  
2) Let $G(z^{-1}) = U(z^{-1})T_{2C}(z^{-1})$; then, let $U(z^{-1}) \in A$ if and only if $G(z^{-1}) \in A$, and
$$G(a_k)\beta_i = 0 \quad \text{for} \ i = 1, \ldots, n_1.$$  

Next, using the concept of bounded linear functional [26], the above conditions will be rewritten. Define
$$\ell(z) = [\beta_1^T \beta_2^T \beta_3^T \ldots] \quad \text{for} \ i = 1, \ldots, n_1 \quad (4.3)$$

where
$$\beta_i = \begin{bmatrix} \beta_{1i} \\ \beta_{2i} \\ \beta_{3i} \\ \vdots \end{bmatrix} \quad \delta_k = (1, a_k, a_k^2, a_k^3, \ldots).$$

Let $y = [y_1 \ y_2]$ where $y_1$ and $y_2$ are, respectively, the inverse Z transform of $G_1(z^{-1})$ and $G_2(z^{-1})$ with $G(z^{-1}) = [G_1(z^{-1}) \ \ G_2(z^{-1})]$. Then, from Theorem A2 (see the Appendix), we have
$$G(a_k)\beta_i = 0 \quad \text{iff} \quad \langle y, \ell(i) \rangle = 0 \quad \text{for} \ i = 1, \ldots, n_1.$$  

Define the subspace
$$B = \{ y \in \ell_1 \mid \langle y, \ell(i) \rangle = 0 \quad \text{for} \ i = 1, \ldots, n_1 \}.$$  

Then, Problem (4.2) can be rewritten as
$$\sigma_0 = \inf_{y \in B} \| \hat{y} - y \|_1. \quad (4.4)$$

To solve Problem (4.4), we need to find the minimum $\sigma_0$ and then construct an $\ell_1$ optimal deconvolution filter to achieve it. The next theorem will show that the calculation of the minimum $\sigma_0$ is equivalent to solving a linear programming problem.

Theorem 1:
1) The minimum $\sigma_0$ can be found by solving the following linear programming problem:
$$\sigma_0 = \max_{\delta_i} \left[ \sum_{i=1}^{m} \delta_i T_1(a_i)\beta_i \right]$$
subject to
$$\max_j \| y_j \|_\infty \leq 1 \quad \text{for} \ j = 1, 2$$

where $y = [y_1 \ y_2] := \sum_{i=1}^{n_1} \delta_i \ell(i)$.

2) There exists an integer $N$ such that
$$\sigma_0 = \max_{\delta_i} \left[ \sum_{i=1}^{m} \delta_i T_1(a_i)\beta_i \right]$$
subject to
$$\max_{j} \max_{k \leq N} |y_j(k)| \leq 1 \quad \text{for} \ j = 1, 2$$

or equivalently
$$\sigma_0 = \max_{\delta_i, \eta} \left[ \sum_{i=1}^{m} \delta_i T_1(a_i)\beta_i \right]$$

subject to
$$|y_j(k)| \leq \eta \quad \text{for} \ j = 1, 2 \quad \text{and} \ k \leq N \quad 0 \leq \eta \leq 1.$$  

Proof:
1) Using Theorem A2, we get
$$\inf_{\delta_i} \| \hat{y} - y \|_1 = \max_{\delta_i} \langle t_1, y \rangle \quad (4.5)$$

where $B^\perp$ is the subspace of $\ell_\infty$ and is spanned by $\ell(i), \ i = 1, \ldots, n_1$. Hence, $y \in B^\perp$ with $\| y \|_\infty \leq 1$ if and only if
$$y = \sum_{i=1}^{n_1} \delta_i \ell(i) \quad (4.6)$$

with
$$\| y \|_\infty = \max_j \| y_j \|_\infty \leq 1.$$  

From Theorem A1 (see the Appendix), we have
$$\langle t_1, y \rangle = \max_{\delta_i} \left[ \sum_{i=1}^{m} \delta_i T_1(a_i)\beta_i \right].$$

2) Since $y_j(k) = \sum_{i=1}^{n_1} \delta_i \beta_i \ell(i)$, where $|a_i| < 1$ for all $i$, it approaches zero as $k$ becomes large. Hence, there exists an integer $N$ such that $|y_j(k)| < \| y \|_\infty$ for all $k < N$. This means that $\| y \|_\infty = \max_{k \leq N} |y_j(k)|$.

3) Obviously, (2) is equivalent to (3). \hfill \square

Remark 2: Since $T_1(z^{-1}) = [L(z^{-1}) \ 0]$, it is obvious that $T_1(a_i)\beta_i = L(a_i)\beta_i$. Define $L(a_i)\beta_i = b_i$ for $i = 1, \ldots, n_1$. The linear programming problem mentioned in Theorem 1 can be further expressed as follows:
$$\sigma_0 = \max_{\delta_i} \sum_{i=1}^{m} \delta_i b_i = \max_{\delta} b^T \delta \quad (4.7)$$

subject to
$$\max_{j} (\| y_j \|_\infty, \| y_j \|_\infty) \leq 1$$

where $b := [b_1 \ \cdots \ b_{n_1}]^T$ and $\delta := [\delta_1 \ \cdots \ \delta_{n_1}]^T$. \hfill \square

By solving the part (3) in Theorem 1, the optimal value $\delta_i$ can be evaluated. Therefore
$$\hat{y} = \sum_{i=1}^{n_1} \delta_i \ell(i) \quad (4.8)$$

achieves the maximum on the right of (4.5) (which will always exist); hence, $\| \hat{y} \|_\infty = 1$. The following theorem is devoted to constructing the optimal solution $\hat{y}$ of Problem (4.4).
Theorem 2: Let \( \hat{g} \) be the optimal solution of (4.5), let \( \hat{s} = t_1 - \hat{g} \), and let \( \check{g} \) be evaluated as (4.8). Then, \( \hat{s} \) satisfies the following conditions:

1) \( \hat{s}(j)\hat{g}(j) \geq 0 \) for \( j = 1, 2 \).
2) \( \hat{s}(j) = 0 \) whenever \( \|\check{g}(j)\| < \max_j \|\check{g}(j)\|_\infty \) for \( j = 1, 2 \).
3) \( \sum_{j=1}^2 \|\hat{s}_j\|_1 = \|\hat{s}\|_1 \) whenever \( \hat{g} \) is not identically equal to zero.
4) \( \|\hat{s}\|_1 = \sigma_0 \).
5) \( \sum_{j=1}^2 \beta_{ij} \sum_{k=0}^{\infty} \hat{s}_j(k)\hat{d}_k \) = \( T_1(a_i)\hat{d}_k \) for \( i = 1, \ldots, n_1 \).

Proof: In view of (4.5) and Theorem A2, \( \hat{g} \in l_\infty \) is aligned with \( \hat{s} \in l_1 \). Conditions (1)–(3) are exactly the same alignment conditions as given in Remark A1 (see the Appendix). In addition, since \( \hat{g} \) is the optimal solution, \( \sigma_0 = \|t_1 - \hat{g}\|_1 = \|\hat{s}\|_1 \), and Condition (4) follows. Recall that \( \hat{g} \in B \); hence

\[
\langle \hat{s}, l(i) \rangle = \langle t_1, l(i) \rangle \quad \text{for } i = 1, \ldots, n_1.
\]

That is

\[
\sum_{j=1}^2 \beta_{ij} \sum_{k=0}^{\infty} \hat{s}_j(k)\hat{d}_k = T_1(a_i)\hat{d}_k \quad \text{for } i = 1, \ldots, n_1. \]

As \( \hat{s} \) is constructed by Theorem 2, the optimal solution \( \hat{g} \) is then computed by \( \hat{g} = t_1 - \hat{s} \). Let \( \hat{G}(z^{-1}) \) be the \( \mathbb{Z} \) transform of \( \hat{g} \). Since \( \hat{G}(z^{-1}) = \hat{F}(z^{-1})T_2(z^{-1}) \in A \), the \( l_1 \) optimal deconvolution filter \( \hat{F}(z^{-1}) \) is also obtained by

\[
\hat{F}(z^{-1}) = \hat{G}(z^{-1})T_2^{-1}(z^{-1})
\]

where \( T_2^{-1}(z^{-1}) \) is the right inverse of \( T_2(z^{-1}) \).

Hereafter, we have discussed the generalized case of solving the \( l_1 \) optimal deconvolution filtering problem under uncertain but bounded inputs and external noises. However, if the system is noise-free or noiseless, the performance index (3.7) is then adopted here for simplicity. Then, (3.7) can be also rewritten as the following one block \( l_1 \) optimization problem:

\[
\sigma_0 = \inf_{F(z^{-1}) \in A} \|T_1(z^{-1}) - F(z^{-1})T_2(z^{-1})\|_A
\]

where \( T_1(z^{-1}) = L(z^{-1}) \), and \( T_2(z^{-1}) = H(z^{-1})L(z^{-1}) \). Let \( G(z^{-1}) = F(z^{-1})T_2(z^{-1}) \in A \). Then, (4.10) is equivalent to

\[
\sigma_0 = \inf_{F(z^{-1}) \in B} \|t_1 - F\|_1
\]

where \( T_1(z^{-1}) = L(z^{-1}) \), and \( T_2(z^{-1}) = H(z^{-1})L(z^{-1}) \). Let \( G(z^{-1}) = F(z^{-1})T_2(z^{-1}) \in A \). Then, (4.10) is equivalent to

\[
\sigma_0 = \inf_{F(z^{-1}) \in B} \|t_1 - F\|_1
\]

To obtain the minimum value \( \sigma_0 \) in (4.11) and to construct an \( l_1 \) optimal deconvolution filter to achieve the minimum \( \sigma_0 \), the following corollaries are presented to achieve these desired results.

Remark 3: Since \( T_1(z^{-1}) \) and \( T_2(z^{-1}) \) are scale functions in (4.10), the solution of \( l_1 \) optimization in (4.11) is more simple. Similar results in control system design are given in [18].

Corollary 1:

1) The minimum \( \sigma_0 \) can be found by solving the following linear programming problem:

\[
\sigma_0 = \max_{t_1} \|t_1 - \hat{g}\|_1
\]

subject to

\[
\sum_{i=1}^{n_1} \delta_i a_i^k \leq 1 \quad k = 0, 1, 2, \ldots.
\]

2) There exists an integer \( N \) such that

\[
\delta_i a_i^k \leq 1 \quad \text{for all } k > N.
\]

That is, only a finite set of constraints is required for solving the linear programming problem.

Proof:

1) Based on Theorem A2, we have

\[
\inf_{F(z^{-1}) \in B} \|t_1 - F\|_1 = \max_{t_1, y} \langle t_1, y \rangle
\]

where \( B^\perp \) is the orthogonal complement of \( B \). Herein, \( B^\perp \) is spanned by \( \delta_i, i = 1, \ldots, n_1 \). Hence, \( y \in B^\perp \) with \( \|y\|_\infty \leq 1 \) if and only if

\[
y = \sum_{i=1}^{n_1} \delta_i \delta_i
\]

with

\[
\|y(k)\| = \sum_{i=1}^{n_1} \delta_i a_i^k \leq 1 \quad \text{for } k = 0, 1, 2, \ldots.
\]

From Theorem A1, we have

\[
\langle t_1, y \rangle = \sum_{i=1}^{n_1} t_1(i) y(k) = \sum_{i=1}^{n_1} \delta_i \sum_{k=0}^{\infty} T_1(a_i)^k
\]

\[
= \sum_{i=1}^{n_1} \delta_i T_1(a_i).
\]

2) Since \( \{y(k)\} = \{\sum_{i=1}^{n_1} \delta_i a_i^k\} \), where \( a_i < 1 \) for all \( i \), it approaches zero as \( k \) becomes large.

By solving the problem in Corollary 1, the optimal value \( \delta_i \)'s are evaluated. This means that

\[
\hat{y} = \sum_{i=1}^{n_1} \delta_i \delta_i
\]

in which \( \delta_i = \{1, a_i^2, \ldots\} \) for \( i = 1, \ldots, n_1 \) can be achieved so that \( \|y\|_\infty = 1 \). The following corollary is devoted to constructing the optimal solution \( \hat{y} \) of (4.11).

Corollary 2:

Let \( \hat{y} \) be the optimal solution of (4.11), and let \( \hat{s} = t_1 - \hat{g} \) and \( \hat{g} \) be evaluated as above. Then, \( \hat{s} \) satisfies the following necessary and sufficient conditions:

1) \( \hat{s}(k)\hat{g}(k) \geq 0 \).
2) \( \hat{s}(k) = 0 \) whenever \( \|\hat{g}(k)\| < 1 \).
3) \( \sum_{k=0}^{\infty} \hat{s}(k) = \sigma_0 \).
4) \( \sum_{k=0}^{\infty} \hat{s}(k) a_i^k = T_1(a_i) \quad i = 1, \ldots, n_1 \).
Proof: From (4.12) and Theorem A2, \( \hat{y} \in \ell_2 \) is aligned with \( \delta \in \ell_2 \). Conditions (1) and (2) are exactly the same alignment conditions as given in Remark A1. Additionally, since \( \hat{y} \) is the optimal solution, \( \sigma_0 = \| \tau - \tilde{y} \|_1 = \| \delta \|_1 \), and then, (3) follows. Recall that \( \tilde{y} \in B \); hence

\[
\langle \tilde{s}, \partial_i \rangle = \langle t_1, \partial_i \rangle \quad i = 1, \cdots, n_1.
\]

That is

\[
\sum_{k=1}^{\infty} \tilde{s}(k)a_k^i = T_1(a_i) \quad i = 1, \cdots, n_1, \quad \square
\]

Since \( B^\perp \) is exactly the subspace of \( \ell_\infty \) spanned by \( \partial_i, \ i = 1, \cdots, n_1 \); therefore, \( B^\perp \subset c_0 \). According to [26], it is well known that the dual space of \( c_0 \), \( c_0^* \), is equal to \( \ell_1 \). If there exists \( Q = B^\perp \), then \( Q^\perp \) is given by

\[
Q^\perp = \{ s \in \ell_1 | \langle \partial_i, s \rangle = 0, \quad i = 1, \cdots, n_1 \}.
\]

Clearly, \( Q^\perp \) is exactly equal to \( B \). Using Theorems A2 and A3 (see the Appendix), it results that

\[
\inf_{g \in B} \| \tau - g \|_1 = \max_{y \in B^\perp} \langle t_1, y \rangle = \max_{y \in Q^\perp} \langle y, t_1 \rangle = \min_{g \in Q^\perp} \| \tau - g \|_1 = \min_{g \in B^\perp} \| \tau - g \|_1.
\]

Thus, it is clear that (4.11) will always have an optimal solution.

When \( \tilde{s} \) is obtained by Corollary 2, the optimal \( \hat{y} \) can be computed by

\[
\hat{y} = t_1 - \tilde{s}.
\]

Let \( \tilde{C}(z^{-1}) \) be the \( Z \)-transform of \( \tilde{s} \). Since \( \hat{C}(z^{-1}) = \tilde{C}(z^{-1})T_2(z^{-1}) \in A \), the optimal deconvolution filter \( \tilde{F}(z^{-1}) \) is then obtained by

\[
\tilde{F}(z^{-1}) = \tilde{C}(z^{-1})T_2^{-1}(z^{-1}). \tag{4.14}
\]

Remark 4:

1) The above analysis is merely restricted to the case of distinct real nonminimum phase zeros of \( T_2(z^{-1}) \). If one of the nonminimum phase zeros \( a_r \) is a complex number, it generates two sequences, namely

\[
R\partial_r = \text{Re} \{ 1, a_r, a_r^2, a_r^3, \cdots \}
\]

and

\[
I\partial_r = \text{Im} \{ 0, a_r, a_r^2, a_r^3, \cdots \}.
\]

2) In the case of \( T_2(z^{-1}) \), there is a zero \( a_r \) with \( m \) multiplicity. On the analogy of Lemma 1, it is seen that \( \hat{F}(z^{-1}) \in A \) if and only if \( G(z^{-1}) \notin A \), and \( G^{(r)}(a_r) = 0 \) for \( r = 0, 1, \cdots, m-1 \), where \( G^{(r)}(z^{-1}) \) denotes the \( r \)th derivative of \( G(z^{-1}) \). Define the sequences \( \xi_i \)'s as follows:

\[
\xi_0 = \{ 1, a_r, a_r^2, a_r^3, \cdots \} = \{ a_r^i \}
\]

\[
\xi_1 = \{ 0, 1, a_r, a_r^2, \cdots \} = \{ \alpha_r^2 \}
\]

\[
\vdots
\]

\[
\xi_{m-1} = \{ \alpha_r^{i-1}(i-1) \cdots (i-m+2) \alpha_r^{m+1} \}
\]

and

\[
B = \{ g \in \ell_1 | \langle g, \xi_i \rangle = 0, \quad \rho = 0, 1, \cdots, m-1 \}.
\]

Then, \( G(z^{-1}) \notin A \) and \( G^{(r)}(a_r) = 0 \) for all \( r \) if and only if \( g \in B \). Note that the total number of sequences will remain finite even if \( T_2(z^{-1}) \) has multiple complex zeros. Hence, the analysis in this section will hold and can be extended.

To summarize the above results, the optimal deconvolution filter design algorithm for (4.1) (or (4.10)) is concluded as follows:

Algorithm:

1) Construct \( T_1(z^{-1}) \) and \( T_2(z^{-1}) \).
2) Perform co-inner–outer factorization to obtain \( T_2(z^{-1}) = T_2CO(z^{-1})T_2CO(z^{-1}) \).
3) Construct a nonzero vector \( \beta_i \) such that \( T_2CO(a_i)\beta_i = 0 \), and \( \ell(i) = \beta_i^T \partial_i \) for \( i = 1, \cdots, n_1 \).
4) Reformulate the problem into an equivalent linear programming problem.
5) Solve the linear programming problem in Theorem 1 (or Corollary 1) to achieve \( \sigma_0 \) and \( \hat{y} \) in (4.8) (or (4.13)).
6) Using (1)–(5) in Theorem 2 (or (1)–(4) in Corollary 2) to construct \( \tilde{s} = \{ \delta_i \} \) (or \( \delta = \{ s[k] \} \)), consequently to \( \tilde{y} = t_1 - \tilde{s} \). And by taking the \( Z \)-transform of \( \tilde{y} \), \( \tilde{G}(z^{-1}) \) is then obtained.
7) The \( \ell_1 \) optimal deconvolution filter embedded into (4.1) (or (4.10)) is given by

\[
\tilde{F}(z^{-1}) = \tilde{C}(z^{-1})T_2^{-1}(z^{-1}).
\]

Steps 1 and 2 involve the implementation of the procedure for co-inner–outer factorization. In Step 4, the problem \( \inf_{g \in B} \| \tau - g \|_1 \) is reformulated into an equivalent linear programming problem by the concept of bounded linear functional, and then, based on Remark 2, the linear programming problem is further transformed into a standard form. Step 5 is to solve the linear programming problem in (4.7) by using simplex algorithm [22] for the reason of its simplicity. When \( T_2(z^{-1}) \) is square (i.e., in the noiseless case), Step 7 would be more simple.

From the above detailed analysis, the properties of the proposed \( \ell_1 \)-optimal deconvolution filtering design algorithm are emphasized in the following remarks.

Remark 5: The complexity of the above design procedure of \( \ell_1 \)-optimal deconvolution filter is heavily dependent on the number of the nonminimum-phase zero of \( T_2(z^{-1}) \) (i.e., \( n_1 \)). However, the design procedure is actually very simple because the actual systems have only a few nonminimum-phase zeros. This will be shown in the example in the following section.

Remark 6: In digital communication, the inputs may be a sequence of pulses with a limited number of discrete value. In general, it is the transmission of a random sequence of +1
and $-1$. When transmitting such data over a communication channel, intersymbol interference may occur. To restore the transmitted signal, the channel must be equalized.

Regardless of modulation and demodulation, an appropriate sampled channel description will fit into our deconvolution structure. In this case, the corresponding equalizer may be obtained by the proposed algorithm in the paper. The output from the equalizer could be fed into a decision module, which determines if the signal $+1$ or $-1$ has been transmitted. Thus, it is evident that the proposed $\ell_1$ deconvolution algorithm can deal with the various bounded amplitude signal sets whether the signal is deterministic or stochastic. It is therefore adequate for use in cases in which the environment has some uncertainties due to driving inputs or external noises.

Remark 7: The design philosophy of the proposed algorithm is to deal with signals having persistent inputs or noises. Such signals cannot be treated in the $\ell_2$ optimization framework due to the infinite energy.

Remark 8: The proposed deconvolution algorithm can also be employed to treat the fixed-lag smoothing problem. Suppose that an estimator has $n$ delays $z^{-n}F(z^{-1})$. This estimator is called a smoother. Then, the deconvolution problem, using the definition of (2.5), is expressed as

$$\sigma = \|[L(z^{-1}) - z^{-n}F(z^{-1})]H(z^{-1})D(z^{-1})\|_A$$

where $T_1(z^{-1}) = [L(z^{-1}) - z^{-n}F(z^{-1})]H(z^{-1})D(z^{-1})$. It is obvious that $T_2(z^{-1}) = [z^{-n}H(z^{-1})L(z^{-1}) - z^{-n}D(z^{-1})]$. Thus, this deconvolution smoothing problem can be solved by combining the method given in Remark 4 with the proposed design algorithm.

V. ILLUSTRATIVE EXAMPLES

To illustrate the design algorithm proposed in Section IV, a nonminimum-phase system with the measurement signal corrupted by colored noise is considered such that

$$y(k) = H(z^{-1})u(k) + D(z^{-1})r(k)$$

where the signal, channel, and noise model are, respectively, specified as follows:

**Signal Model:**

$$L(z^{-1}) = \frac{0.43 + 0.4085z^{-1}}{1 - 0.1584z^{-1}}$$

**Channel Model:**

$$H(z^{-1}) = \frac{0.9 + z^{-1}}{1 + 0.1z^{-1}}$$

**Noise Model:**

$$D(z^{-1}) = \frac{0.043 + 0.0409z^{-1}}{1 - 0.1584z^{-1}}.$$

Then, according to the proposed design algorithm, the corresponding $\ell_1$ optimal deconvolution filter $\hat{F}(z^{-1})$ can be easily obtained by the following steps as the input $r(k)$ and external noise $\nu(k)$ are uncertain but bounded:

**Step 1)** Construct

$$T_1(z^{-1}) = [L(z^{-1}) - z^{-n}F(z^{-1})]H(z^{-1})D(z^{-1})$$

$$T_2(z^{-1}) = [H(z^{-1})L(z^{-1}) - z^{-n}D(z^{-1})]$$

**Step 2)** Perform co-inner–outer factorization to get

$$T_2(z^{-1}) = T_{2CO}(z^{-1})T_{2CI}(z^{-1})$$

**Step 3)** The nonminimum phase zero of $T_{2CI}(z^{-1})$ is only at $z = -1.1111$. In this case, $n_1 = 1$. Construct a nonzero vector $\beta = [1 \ 0]^T$ such that $T_{2CI}(-0.9)\beta = 0$. Furthermore, construct

$$\ell(1) = \beta^T \partial = [\partial \ 0]$$

where

$$\partial = (1, -0.9, 0.81, \cdots).$$

**Step 4)** Reformulate the problem to an equivalent linear programming problem, i.e.

$$\sigma_0 = \max_\delta \delta T_1(-0.9)\beta = \max_\delta \delta \cdot 0.05457$$

subject to

$$\|y\|_\infty = \|\delta \ell(1)\|_\infty \leq 1.$$
Step 7) Since $T_1(\hat{z}^{-1}) - \hat{F}(\hat{z}^{-1})T_2(\hat{z}^{-1}) = \hat{S}(\hat{z}^{-1})$, the optimal deconvolution filter $\hat{F}(\hat{z}^{-1})$ is easily computed as

$$\hat{F}(\hat{z}^{-1}) = \frac{0.3754 + 0.7133\hat{z}^{-1} + 0.3716\hat{z}^{-2} + 0.0004\hat{z}^{-3}}{0.4343 + 1.1874\hat{z}^{-1} + 1.0645\hat{z}^{-2} + 0.3300\hat{z}^{-3}}.$$  

For explaining the robustness to the proposed algorithm to the uncertain signal characteristics and parameter variations, an optimal deconvolution filter is obtained by the following minimum mean square error (MMSE) criterion:

$$\min_{F} J = \min_{F} \|L(1 - FH) : -FD\|_2.$$  

By the inner–outer factorization method in [25], an $\ell_2$ equalizer is found as

$$F_{\text{MSE}}(\hat{z}^{-1}) = \frac{1.1099 + 0.9287\hat{z}^{-1} + 0.0818\hat{z}^{-2}}{1 + 1.818\hat{z}^{-1} + 0.3234\hat{z}^{-2}}.$$  

Assume that the driving signal $\{r(k)\}$ and the external noise $\{v(k)\}$ are two independent white and zero mean with variances $\lambda_r^2 = 1$ and $\lambda_v^2$, respectively, equal to 1, 5, 10, and 20. Let a sequence of input signal be transmitted to the nonminimum phase channel. The transmitted signal $\{u(k)\}$ and the estimates $\{\hat{u}_1(k)\}$ and $\{\hat{u}_2(k)\}$ obtained by, respectively, passing $\{y(k)\}$ through the $\hat{F}(\hat{z}^{-1})$ and $F_{\text{MSE}}(\hat{z}^{-1})$ are shown in Fig. 2(a)–(d). From Fig. 2, under small variances of noise, we find that the estimate $\{\hat{u}_2(k)\}$ based on the quadratic integral type criterion is better than that obtained by the proposed algorithm. However, when variances of noise are large, the estimate based on the quadratic integral type criterion will be less robust than that obtained by the proposed algorithm.

However, the performance may be degraded by the uncertainties due to the variations of the statistical properties of the driving signal and the external noise. In order to determine if the estimate based on the quadratic integral type criterion remains better than that obtained by the proposed algorithm, $\hat{F}(\hat{z}^{-1})$ and $F_{\text{MSE}}(\hat{z}^{-1})$, respectively, operate on the four uncertain cases in input signals and noises shown in Table I, where $N(0, \lambda_r^2)$ and $U(0, \lambda_v^2)$ denote the normal distribution and uniform distribution, respectively.
The corresponding simulation results are given in Figs. 3–6. From Fig. 3, we find that in Case 1, the value of \( \lambda_2^2 \) varies within 2.4843–12.4511, and the Mean Square Error (MSE) caused by \( f_{\text{MSE}}(z^{-1}) \) is smaller than that caused by \( \hat{F}(z^{-1}) \). When the value of \( \lambda_2^2 \) is larger than 12.4511, the MSE error caused by using \( f_{\text{MSE}}(z^{-1}) \) increases at a faster rate than that by \( \hat{F}(z^{-1}) \). In addition, from Figs. 4–6, it is found that the deconvolution filter designed by the proposed algorithm is more insensitive to the variations of \( \lambda_2^2 \) than that designed by the \( \ell_2 \) algorithm in [9] when the external noises in Cases 2–4 are of non-Gaussian distribution. It is clear that the proposed algorithm is not only adequate for dealing with cases of signals with infinite energy but is also more robust for cases when the system suffers from uncertainties caused by variations of the statistical properties of inputs and external noises.

VI. CONCLUSIONS

In this paper, we have studied the deconvolution filtering design problem for the systems with uncertain-but-bounded inputs and external noises. A stable and causal \( \ell_2 \) optimal deconvolution filter is designed to minimize the maximum peak gain of the output error from the viewpoint of time domain. The solution consists of two parts: first, the calculation of minimum value via an equivalent A-norm minimization and, second, construction of the \( \ell_2 \) optimal deconvolution filter by solving a set of constrained linear equations. The existence of the \( \ell_2 \) optimal deconvolution filter is discussed, and an equivalent linear programming problem is formulated to numerically solve the problem. A summary design algorithm is also proposed for the implementation of \( \ell_1 \) optimal deconvolution filter. The complexity of the design procedure is heavily dependent on the number of nonminimum phase zeros of the channel system. The proposed solution suggests a design philosophy to treat the deconvolution problem from the viewpoint of the time domain when inputs and external noises are not clearly specified. The calculation of the proposed algorithm is simple. Finally, one numerical example has been given to illustrate the noticeable advantages of the proposed algorithm. Several simulation results have exhibited strong robustness against the uncertainties of both driving signal and corrupted noise.

APPENDIX

The following theorems that play a major role in the analysis are given below. The detailed proof can be found in [26].

**Theorem A1** [26]: Every bounded linear functional on \( \ell_1 \) is uniquely represented in the form

\[
\langle x, w \rangle = \sum_{i=1}^{n} \sum_{k=0}^{\infty} w_i(k)x_i(k)
\]

where \( x \in \ell_1 \) and \( w \in \ell_\infty \). Furthermore, every element of \( \ell_\infty \) defines a member of \((\ell_1)^*\) in this way, and we have

\[
\|w\|_\infty = \sup_{x \neq 0} \frac{|\langle x, w \rangle|}{\|x\|_1}.
\]

\( \square \)
Remark A1: Theorem A1 proves that the dual space of $\ell_1$ is $\ell_\infty$. An immediate consequence of this theorem is the following result: If $w \in \ell_\infty$ is aligned with $x \in \ell_1$, i.e., $\langle x, w \rangle = \|x\|_1 \cdot \|w\|_\infty$, then it is obvious that $w$ and $x$ satisfy the following conditions:

1) $w(k)x(k) \geq 0$.
2) $x_i = 0$ whenever $|w_i| < \|w\|_\infty$, or $|w(k)| = \|w\|_\infty$ whenever $x_i(k) \neq 0$.
3) $\sum_{i=1}^{\infty} |x_i|_1 = \|x\|_1$ whenever $w$ is not identically equal to zero.

Remark A2: The dual space of $\ell_\infty$ is not $\ell_1$. Let $c_0$ denote the subspace of $\ell_\infty$ consisting of all sequences that converge to zero. Then, it can be shown that the dual space of $c_0$ is $\ell_1$ [26], where the linear functional are defined as in the above theorem.

Theorem A2 [26]: Let $x \in X$ and $\mu$ denote its distance from the subspace $B$, $B \subset X$. Then

$$\mu = \inf_{b \in B} \|x - b\|_p = \max_{x^* \in B^*} \langle x, x^* \rangle$$

where the maximum on the right is achieved for some $x^* \in B^*$, with $\|x^*\|_p = 1$. If the infimum on the left is achieved for some $b \in B$, then $x^*$ is aligned with $x - b$.

Theorem A3 [26]: Let $B$ be a subspace of $X$ and $x^* \in X^*$ be a distance $\mu$ from $B$. Then

$$\mu = \min_{b \in B} \|x^* - b\|_q = \sup_{x^* \in B} \langle x^*, x^* \rangle$$

where the minimum on the right is achieved for some $\tilde{b} \in B$. If the supremum on the right is achieved for some $\tilde{x} \in X$, then $\|\tilde{x}\|_p = 1$, and $x^* = \tilde{x}$ is aligned with $\tilde{x}$.

Remark A3: The above two theorems introduce a duality principle stating the equivalence of two extremization problems: one formulated in a normed space and the other in its dual.

REFERENCES