Available online at www.sciencedirect.com
SCIENCE DIRECT®
Systems & Control Letters 54 (2005) 125–134
SYSTEMS & CONTROL LETTERS
www.elsevier.com/locate/sysconle

# On the algebraic identifiability of finite impulse response channels driven by linearly precoded signals

Jonathan H. Manton[a,*], Walter D. Neumann[b], Paul T. Norbury[c]

[a]*ARC Special Research Centre for Ultra-BroadBand Information Networks, Department of Electrical and Electronic Engineering, The University of Melbourne, Melbourne, Victoria 3010, Australia*
[b]*Department of Mathematics, Columbia University, New York, NY 10027, USA*
[c]*Department of Mathematics, The University of Melbourne, Melbourne, Victoria 3010, Australia*

Received 1 November 2001; accepted 16 July 2004
Available online 9 September 2004

**Abstract**

It is common in wireless communications to perform some form of linear precoding operation on the source signal prior to transmission over a channel. Although the traditional reason for precoding is to improve the performance of the communication system, this paper draws attention to the fact that the receiver can identify the impulse response of the channel without any prior knowledge of the transmitted signal simply by solving a system of polynomial equations. Since different precoders lead to different systems of equations, this paper addresses the fundamental issue of determining which classes of linear precoders lead to a system of equations having a unique solution. In doing so, basic properties of polynomial equations which are useful for studying other identifiability issues commonly encountered in engineering and the applied sciences are presented.


*Keywords:* Linear precoders; Wireless communications; Polynomial equations; Blind identification

## 1. Introduction

The need often arises for the estimation of the impulse response $\{h_0, \ldots, h_l\}$ of the noise-corrupted convolutive channel

$$y_i = \sum_{k=0}^{l} h_k x_{i-k} + w_i, \quad i = \ldots, -1, 0, 1, \ldots, \qquad (1)$$

where $\{x_i\}_{i=-\infty}^{\infty}$ is the transmitted signal, $\{y_i\}_{i=-\infty}^{\infty}$ the received signal and $\{w_i\}_{i=-\infty}^{\infty}$ an unobserved noise signal, typically assumed to be white Gaussian. Since the transmitted signal is unknown to the receiver, this channel identification problem can only be solved if the receiver knows some property about the transmitted signal, such as a statistical property [7,18] or finite alphabet property [5,9,19] of $\{x_i\}_{i=-\infty}^{\infty}$. This paper studies a new method of identifying the channel based on prior knowledge of the algebraic structure of $\{x_i\}_{i=-\infty}^{\infty}$.

* Corresponding author. Tel.: +61-3-83446791; fax: +61-3-83446678.
*E-mail address:* jon@ee.mu.oz.au (J.H. Manton).

0167-6911/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.sysconle.2004.07.004

The motivation for this work is that the use of linear precoders in wireless transmission systems is quite common; see [8,11–13,15,20] and the references therein. As noted in [10,15,20], linear precoders introduce an algebraic structure which can be exploited at the receiver to estimate (or refine a previously obtained estimate of) the channel. Estimating the channel in this way is called algebraic channel identification because no statistical, finite alphabet or any other properties besides the algebraic structure of the transmitted signal are assumed.

*The main contribution of this paper is to further the understanding of what classes of linear precoders introduce sufficient algebraic structure to make it feasible for the receiver to identify the channel.*

This paper differs from others in two important respects. Often the main mathematical tools used to study precoders are linear algebra and $z$-domain analysis, hence only a restricted class of precoders can be studied [20]. The present paper uses results from algebraic geometry to study arbitrary linear precoders. Moreover, the definition of identifiability is often based on whether or not a specific algorithm can identify the channel whereas the present paper takes the definition to be whether or not it is theoretically possible to identify the channel.

As will be shown in subsequent sections, the algebraic channel identifiability problem reduces to determining if a system of polynomial equations is invertible. This fundamental problem is known to be non-trivial [17]. Therefore, *the secondary contribution of this paper is to review a number of results from algebraic geometry which facilitate the determination of the invertibility of a system of polynomial equations.*

The paper is organised as follows. Section 2 states the algebraic identifiability problem. Section 3 sketches fundamental properties of polynomial maps required in later sections. Section 4 derives the first main result, Theorem 7, which states the original identifiability problem (complete with additive noise and scale ambiguity) can be reduced, without any loss of generality, to a significantly simpler identifiability problem (one without ambiguity and without noise). The second main result, Proposition 8 and Theorem 9, is presented in Section 5. It is proved that, with the exception of some non-generic precoders, the ability of a precoder to enable the receiver to identify the channel depends only on the size of the precoder matrix and not on its elements. Section 6 concludes with a summary of the main results.

## 2. System model

The linearly precoded communication system considered in this paper is as follows. Assume $p$ complex valued source symbols are transmitted over a finite impulse response channel of order $l$ by first linearly precoding them to form $n+l$ symbols. Let $\boldsymbol{s} \in \mathbb{C}^p$ denote the $p$ source symbols and $\boldsymbol{x} = [x_{1-l}, \ldots, x_n]^{\mathrm{T}} \in \mathbb{C}^{n+l}$ the linearly precoded symbols; by definition, $\boldsymbol{x} = P\boldsymbol{s}$ for some precoder matrix $P \in \mathbb{C}^{(n+l)\times p}$. If $\boldsymbol{x}$ is transmitted through a channel whose impulse response is $\tilde{\boldsymbol{h}} = [h_0, \ldots, h_l]^{\mathrm{T}} \in \mathbb{C}^{l+1}$ then the received vector $\boldsymbol{y} = [y_1, \ldots, y_n]^{\mathrm{T}} \in \mathbb{C}^n$ of length $n$ is related to $\boldsymbol{x}$ by the convolution

$$y_i = \sum_{k=0}^{l} h_k x_{i-k} + w_i, \quad i = 1, \ldots, n, \tag{2}$$

where $w_i$ denotes a sequence of random variables modelling the additive noise. The actual distribution of the noise is relatively unimportant in this paper; it suffices for the distribution to be absolutely continuous with respect to Lebesgue measure (see Section 3.3). Note that Gaussian noise satisfies this condition.

**Remark.** The reason for the tilde is that the vector $\boldsymbol{h} = [h_1, \ldots, h_l]^{\mathrm{T}}$ is introduced in Section 4 since it suffices later to assume that $h_0 = 1$ in (2).

It is assumed throughout that the precoder matrix $P$ has full column rank and that $n \geqslant p$. This is always the case in practice for otherwise the source vector $\boldsymbol{s}$ could not be recovered from the output $\boldsymbol{y}$ even if the channel $\tilde{\boldsymbol{h}}$ were known.

In the literature [18], (2) is often written in matrix form as $\boldsymbol{y} = \tilde{H}\boldsymbol{x} + \boldsymbol{w} = \tilde{H}P\boldsymbol{s} + \boldsymbol{w}$ where $\tilde{H} \in \mathbb{C}^{n\times(n+l)}$ is the upper triangular Toeplitz matrix having $[h_l, h_{l-1}, \ldots, h_0, 0, \ldots, 0]$ as its first row.

Since it is assumed that only $\boldsymbol{y}$ is known to the receiver, the equation $\boldsymbol{y} = \tilde{H}P\boldsymbol{s} + \boldsymbol{w}$ is not a linear equation but a bilinear equation in $\tilde{\boldsymbol{h}}$ and $\boldsymbol{s}$. To make this distinction explicit, the function $\tilde{F}(\boldsymbol{s}, \tilde{\boldsymbol{h}}) = \tilde{H}P\boldsymbol{s}$

is introduced, so that

$$\boldsymbol{y} = \tilde{F}(\boldsymbol{s}, \tilde{\boldsymbol{h}}) + \boldsymbol{w}. \tag{3}$$

Given only $\boldsymbol{y}$, the channel $\tilde{\boldsymbol{h}}$ can be estimated in the least-squares sense by finding an $\hat{\tilde{\boldsymbol{h}}}$ which minimises $\inf_{\hat{\boldsymbol{s}}} \|\boldsymbol{y} - \tilde{F}(\hat{\boldsymbol{s}}, \hat{\tilde{\boldsymbol{h}}})\|^2$; an algorithm for doing so is given in [10]. Roughly speaking, the channel will be said to be algebraically identifiable if the least-squares estimate is unique. However, there are two refinements that must be made to this definition. The first refinement is to allow for an unknown scaling factor in the channel estimate; if $\hat{\tilde{\boldsymbol{h}}}$ is a least-squares estimate then so too is $\lambda\hat{\tilde{\boldsymbol{h}}}$ for any non-zero $\lambda \in \mathbb{C}$, a consequence of the equality $\tilde{F}(\lambda \boldsymbol{s}, \lambda^{-1}\tilde{\boldsymbol{h}}) = \tilde{F}(\boldsymbol{s}, \tilde{\boldsymbol{h}})$. The second refinement is to take into account the dependence on the actual symbols transmitted. For instance, if $\boldsymbol{s}$ is the zero vector then it is impossible to identify the channel. These refinements are made in Section 4 since the results of Section 3 are required.

## 3. Polynomial maps

Since $\tilde{F}$ in (3) is a bilinear map, and hence a polynomial map, the study of such maps is the key to understanding the intricacies of channel identification.

Although the results in this section are known in the algebraic geometry community and are thus stated here without proof, it is not easy to find explicit statements of them in the literature. The reader is referred to [4,6,14,16] for more information.

Throughout, $\|\cdot\|$ denotes the Euclidean norm and $B(\boldsymbol{z}; r)$ denotes the open ball centred at $\boldsymbol{z}$ with radius $r$. Topological concepts such as openness and denseness are with respect to the usual topology (and not with respect to the Zariski topology [3] often used in algebraic geometry).

### 3.1. Generic number of pre-images and generic points

Let $F : \mathbb{C}^m \to \mathbb{C}^n$ be a polynomial map. The pre-images of a point $\boldsymbol{y} \in \mathbb{C}^n$ under $F$ are the elements of the set $\{\boldsymbol{z} \in \mathbb{C}^m : \boldsymbol{y} = F(\boldsymbol{z})\}$. Theorem 1 states that, for most $\boldsymbol{y}$, the number of pre-images is constant, and moreover, that the pre-images behave in a predictable way if $\boldsymbol{y}$ is perturbed slightly.

**Theorem 1.** *Let $F : \mathbb{C}^m \to \mathbb{C}^n$ be a polynomial map and let $V = \overline{F(\mathbb{C}^m)} \subset \mathbb{C}^n$ denote the closure of its image. There exists a unique number $N$, possibly infinite, and a polynomial $h$: $\mathbb{C}^n \to \mathbb{C}$ used to define the set $W = \{\boldsymbol{y} \in V : h(\boldsymbol{y}) \neq 0\}$, which satisfies the following.* (1) *The set $W$ is open and dense in $V$.* (2) *For any $\boldsymbol{y} \in W$, there are precisely $N$ pre-images of $\boldsymbol{y}$ under $F$.* (3) *If $N$ is finite, let $\boldsymbol{z}_1 \in \mathbb{C}^m$ be such that $F(\boldsymbol{z}_1) \in W$ and let $\boldsymbol{z}_2, \ldots, \boldsymbol{z}_N$ be the other $N-1$ pre-images, that is, $F(\boldsymbol{z}_1) = \cdots = F(\boldsymbol{z}_N)$. Then for any $\varepsilon > 0$ there exists a $\delta > 0$ such that for any $\boldsymbol{z} \in B(\boldsymbol{z}_1; \delta)$ there are precisely $N$ pre-images of $F(\boldsymbol{z})$, and moreover, each set $B(\boldsymbol{z}_i; \varepsilon)$ contains a pre-image of $F(\boldsymbol{z})$ for $i = 1, \ldots, N$.*

**Remark.** Part 3 of Theorem 1 can be replaced by the stronger requirement that the restriction of $F$ to the set $\{\boldsymbol{z} : F(\boldsymbol{z}) \in W\}$ is a covering map; see [1, Section 10.4] for the definition of a covering map.

Pre-images of $F(\boldsymbol{z})$ under $F$ for a given point $z \in \mathbb{C}^m$ feature prominently in this paper. Define $N$ and $h$ as in Theorem 1; it follows that the polynomial $g$: $\mathbb{C}^m \to \mathbb{C}$ defined by $g(\boldsymbol{z}) = h(F(\boldsymbol{z}))$ is not the zero polynomial, and moreover, if $g(\boldsymbol{z}) \neq 0$ then $F(\boldsymbol{z})$ has precisely $N$ pre-images. This motivates the following definitions.

A property which holds for all $\boldsymbol{z}$ for which $g(\boldsymbol{z}) \neq 0$ for some non-zero polynomial $g$ is said to hold for *generic* $\boldsymbol{z}$; note that $\{\boldsymbol{z} \in \mathbb{C}^m : g(\boldsymbol{z}) \neq 0\}$ is open and dense in $\mathbb{C}^m$. The number $N$ in Theorem 1 is called the *generic number of pre-images* of $F$. Any $\boldsymbol{z}$ for which there exists a $W$ defined as in Theorem 1 and such that $F(\boldsymbol{z}) \in W$ is called a *generic point* of $F$. A generic point should be thought of as a well-behaved point in the sense that part 3 of Theorem 1 holds about a generic point.

Sometimes, it is convenient to partition $\boldsymbol{z}$ as $\boldsymbol{z} = (\boldsymbol{z}_1, \boldsymbol{z}_2)$. If a property holds for generic $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ then it clearly holds for generic $\boldsymbol{z}$. Lemma 2 states the converse; if a property holds for generic $\boldsymbol{z}$ then, for a generic $\boldsymbol{z}_1$, it holds for generic $\boldsymbol{z}_2$ (and, for a generic $\boldsymbol{z}_2$, it holds for generic $\boldsymbol{z}_1$).

**Lemma 2.** *Let $g$: $\mathbb{C}^m \times \mathbb{C}^n \to \mathbb{C}$ be a non-zero polynomial. There exists a non-zero polynomial $g_1$: $\mathbb{C}^m \to \mathbb{C}$ such that, for all $\boldsymbol{z}_1 \in \mathbb{C}^m$ satisfying $g_1(\boldsymbol{z}_1) \neq 0$, there exists a non-zero polynomial*

$g_2$: $\mathbb{C}^n \to \mathbb{C}$ (*possibly depending on* $\boldsymbol{z}_1$) *such that* $g_2(\boldsymbol{z}_2) \neq 0$ *implies* $g(\boldsymbol{z}_1, \boldsymbol{z}_2) \neq 0$.

### 3.2. Invertible, rationally invertible and full rank maps

A polynomial map $F$ is *invertible* if, for all $\boldsymbol{z}$, $F(\boldsymbol{z})$ has one pre-image. It is *rationally invertible* if $F(\boldsymbol{z})$ has one pre-image for generic $\boldsymbol{z}$. It has *full rank* if the generic number of pre-images of $F(\boldsymbol{z})$ is finite. It is known that $F$ is invertible if and only if there exists a polynomial map $G$ such that $G \circ F$, the composition of $G$ and $F$, is the identity map. Similarly, $F$ is rationally invertible if and only if there exists a rational function $G$ such that $G \circ F$ is the identity map over its domain of definition [3,14]. Proposition 3 below justifies the term full rank; it requires the following definitions.

Let $f_1, \ldots, f_n$ be polynomials from $\mathbb{C}^m$ to $\mathbb{C}$ and define the polynomial map $F$: $\mathbb{C}^m \to \mathbb{C}^n$ so that $F(\boldsymbol{z}) = (f_1(\boldsymbol{z}), \ldots, f_n(\boldsymbol{z})) \in \mathbb{C}^n$ for $\boldsymbol{z} \in \mathbb{C}^m$. The *Jacobian matrix* $J$ of $F$ is a polynomial matrix whose $ij$th entry is the polynomial $\partial f_i / \partial \boldsymbol{z}_j$. It has a well-defined rank as a matrix over the ring of polynomials.

**Proposition 3.** *Let $J$ be the Jacobian matrix of the polynomial map $F$. The generic number of pre-images of $F$ is finite if and only if $J$ has full column rank.*

**Remark.** Proposition 3 is a consequence of the stronger result that the rank of $J$ equals the largest number of algebraically independent polynomials in the set $\{f_1, \ldots, f_n\}$; see [14].

The evaluation of the Jacobian matrix at a point $\boldsymbol{z}$ is denoted by $J_{\boldsymbol{z}}$ and is a complex valued matrix. The rank of $J_{\boldsymbol{z}}$ never exceeds the rank of $J$, and for generic $\boldsymbol{z}$, the ranks are equal. Thus, $F$ has full rank if and only if there exists a $\boldsymbol{z}$ such that $J_{\boldsymbol{z}}$ has full column rank. Another practical test is the following.

**Proposition 4.** *Let $F$: $\mathbb{C}^m \to \mathbb{C}^n$ be a polynomial map and assume there exists a point $\boldsymbol{y} \in \mathbb{C}^n$ such that the number of pre-images $N$ of $\boldsymbol{y}$ under $F$ satisfies $1 \leqslant N < \infty$. Then $F$ has full rank, and moreover, if $m = n$ then the generic number of pre-images of $F$ is greater than or equal to $N$.*

Unfortunately, testing for rational invertibility is harder [14,17] because it does not suffice to show that there is a single pre-image of $F(\boldsymbol{z})$ for a particular $\boldsymbol{z}$. One complication is the possibility of a pre-image hiding at infinity.[1] Later, in Proposition 12, a sufficient condition for a map to be rationally invertible is derived.

**Remark.** Testing for invertibility is also non-trivial; it has long been conjectured that a polynomial map $F$: $\mathbb{C}^n \to \mathbb{C}^n$ is invertible if and only if the determinant of its Jacobian matrix is a non-zero constant [2].

### 3.3. Randomness and least-squares solutions

In the presence of additive noise $\boldsymbol{w}$, the polynomial equation $\boldsymbol{y} = F(\boldsymbol{z}) + \boldsymbol{w}$ is often solved in the least-squares sense: find $\hat{\boldsymbol{z}}$ to minimise $\|\boldsymbol{y} - F(\hat{z})\|^2$. A fundamental question is whether or not there is a unique minimum. Proposition 5 implies that provided $\boldsymbol{w}$ is chosen at random, the number of minimising points is equal to the generic number of pre-images of $F$ almost surely. As will be seen later, this means that it suffices to ignore the additive noise when studying identifiability issues.

A vector $\boldsymbol{w} \in \mathbb{C}^n$ is *chosen at random* if the vector $[\Re\boldsymbol{w}^{\mathrm{T}}\ \Im\boldsymbol{w}^{\mathrm{T}}]^{\mathrm{T}}$ formed from the real and imaginary components of $\boldsymbol{w}$ is a realisation of a $2n$-dimensional real random vector whose probability distribution is absolutely continuous with respect to $2n$-dimensional Lebesgue measure.

**Proposition 5.** *Let $F$: $\mathbb{C}^m \to \mathbb{C}^n$ be a polynomial map, let $\boldsymbol{z} \in \mathbb{C}^m$ be chosen arbitrarily and let $\boldsymbol{w} \in \mathbb{C}^n$ be chosen at random. Define $\Theta \subset \mathbb{C}^m$ to be the set of all points $\hat{\boldsymbol{z}} \in \mathbb{C}^m$ which minimise $\|F(\boldsymbol{z}) + \boldsymbol{w} - F(\hat{\boldsymbol{z}})\|^2$. Then, with probability one, there exists a $\bar{\boldsymbol{y}}$ in the image of $F$ such that $\Theta = \{\hat{\boldsymbol{z}} \in \mathbb{C}^m : \bar{\boldsymbol{y}} = F(\hat{\boldsymbol{z}})\}$. Furthermore, for any polynomial $h$: $\mathbb{C}^n \to \mathbb{C}$ which is not identically zero on the image of $F$, $h(\bar{\boldsymbol{y}}) \neq 0$ with probability one.*

---

[1] By definition, $F(\boldsymbol{z})$ has a pre-image at infinity if there is a sequence $\{\boldsymbol{z}_k\}_{k=1}^{\infty}$ which diverges ($\|\boldsymbol{z}_k\| \to \infty$) yet $F(\boldsymbol{z}_k) \to F(\boldsymbol{z})$. For instance, $F(\boldsymbol{z}_1, \boldsymbol{z}_2) = (\boldsymbol{z}_1\boldsymbol{z}_2, \boldsymbol{z}_2(\boldsymbol{z}_2 - 1))$ generically has two pre-images but $F(\boldsymbol{z}_1, \boldsymbol{z}_2) = (1, 0)$ has only one solution. The other solution is hiding at infinity; $F(k, 1/k) = (1, 1/k(1/k - 1)) \to (1, 0)$.

**Remark.** In fact, the $\bar{\boldsymbol{y}}$ in Proposition 5 is the Euclidean projection of $\boldsymbol{y}$ onto theclosure of the image of $F$.

## 4. Algebraic channel identifiability

This section uses the results of the previous section to define precisely the concept of algebraic identifiability introduced in Section 2. A technique for converting the original identifiability problem into a simpler one without scale ambiguity or additive noise is given.

As in Section 2, let $\boldsymbol{y} = \tilde{F}(\boldsymbol{s}, \tilde{\boldsymbol{h}}) + \boldsymbol{w}$ be the received vector. For a given $\boldsymbol{y}$, define $\Theta \subset \mathbb{C}^{l+1}$ to be the set of all least-squares estimates of the channel. That is, $\hat{\tilde{\boldsymbol{h}}} \in \Theta$ if and only if there exists an $\hat{\boldsymbol{s}} \in \mathbb{C}^p$ such that $(\hat{\boldsymbol{s}}, \hat{\tilde{\boldsymbol{h}}})$ is a global minimum of the cost function

$$\|\boldsymbol{y} - \tilde{F}(\hat{\boldsymbol{s}}, \hat{\tilde{\boldsymbol{h}}})\|^2. \tag{4}$$

Note that uniqueness of $\hat{\boldsymbol{s}}$ is not required.

Since $\hat{\tilde{\boldsymbol{h}}} \in \Theta$ implies $\lambda\hat{\tilde{\boldsymbol{h}}} \in \Theta$ for any non-zero $\lambda \in \mathbb{C}$ (see Section 2), it is more interesting to study equivalence classes of $\Theta$, where $\tilde{\boldsymbol{h}}_1, \tilde{\boldsymbol{h}}_2 \in \Theta$ are equivalent if there exists a non-zero $\lambda \in \mathbb{C}$ such that $\tilde{\boldsymbol{h}}_1 = \lambda\tilde{\boldsymbol{h}}_2$. (Such equivalence classes of $\mathbb{C}^{l+1}$ with the origin omitted form $l$-dimensional complex projective space [3,10].) Define $N(\boldsymbol{s}, \tilde{\boldsymbol{h}}, \boldsymbol{w})$ to be the number of elements up to equivalence of $\Theta$; here the dependence of $\Theta$ on $\boldsymbol{y}$, and hence on $\boldsymbol{s}$, $\tilde{\boldsymbol{h}}$ and $\boldsymbol{w}$, is made explicit.

Observe that, for a given source vector $\boldsymbol{s}$, channel vector $\tilde{\boldsymbol{h}}$ and noise vector $\boldsymbol{w}$, if $N(\boldsymbol{s}, \tilde{\boldsymbol{h}}, \boldsymbol{w}) = 1$ then the receiver can identify the channel $\tilde{\boldsymbol{h}}$ uniquely up to an unknown complex valued scaling factor based only on the received vector $\boldsymbol{y} = \tilde{F}(\boldsymbol{s}, \tilde{\boldsymbol{h}}) + \boldsymbol{w}$ by computing the least-squares estimate of $\tilde{\boldsymbol{h}}$. It is important then to study how $N(\boldsymbol{s}, \tilde{\boldsymbol{h}}, \boldsymbol{w})$ depends on $\boldsymbol{s}$, $\tilde{\boldsymbol{h}}$ and $\boldsymbol{w}$. Theorem 7 below proves that $N(\boldsymbol{s}, \tilde{\boldsymbol{h}}, \boldsymbol{w})$ is constant for almost all $\boldsymbol{s}$, $\tilde{\boldsymbol{h}}$ and $\boldsymbol{w}$, and moreover, that this constant value can be determined by studying a related but simpler identifiability problem.

The statement of the theorem requires the polynomial map $F\colon \mathbb{C}^{p+l} \to \mathbb{C}^n$ defined as

$$F(\boldsymbol{s}, \boldsymbol{h}) = \tilde{F}(\boldsymbol{s}, [1\ \ \boldsymbol{h}^{\mathrm{T}}]^{\mathrm{T}}) = HP\boldsymbol{s},$$
$$\boldsymbol{s} \in \mathbb{C}^p, \quad \boldsymbol{h} = [h_1, \ldots, h_l]^{\mathrm{T}} \in \mathbb{C}^l, \tag{5}$$

where $H \in \mathbb{C}^{n\times(n+l)}$ is the upper triangular Toeplitz matrix with first row $[h_l, h_{l-1}, \ldots, h_1, 1, 0, \ldots, 0]$. Note that $\boldsymbol{y} = F(\boldsymbol{s}, \boldsymbol{h})$ is simply (2) with $h_0 = 1$.

To motivate the theorem, consider the four sets

$$X_1(\boldsymbol{y}) = \{(\boldsymbol{s}, \tilde{\boldsymbol{h}})\colon \tilde{F}(\boldsymbol{s}, \tilde{\boldsymbol{h}}) = \boldsymbol{y}\},$$
$$X_2(\boldsymbol{y}) = \{\tilde{\boldsymbol{h}}\colon \exists \boldsymbol{s}, \tilde{F}(\boldsymbol{s}, \tilde{\boldsymbol{h}}) = \boldsymbol{y}\}, \tag{6}$$

$$X_3(\boldsymbol{y}) = \{(\boldsymbol{s}, \boldsymbol{h})\colon F(\boldsymbol{s}, \boldsymbol{h}) = \boldsymbol{y}\},$$
$$X_4(\boldsymbol{y}) = \{\boldsymbol{h}\colon \exists \boldsymbol{s}, F(\boldsymbol{s}, \boldsymbol{h}) = \boldsymbol{y}\}, \tag{7}$$

where $\boldsymbol{s} \in \mathbb{C}^p$, $\tilde{\boldsymbol{h}} \in \mathbb{C}^{l+1}$ and $\boldsymbol{h} \in \mathbb{C}^l$. Observe that, in the noise free case ($\boldsymbol{w} = 0$), $X_2(\boldsymbol{y})$ equals the set of least-squares channel estimates $\Theta$. When noise is present then Proposition 5 implies that, with probability one, there exists a $\bar{\boldsymbol{y}}$ such that the set of global minima of (4) equals $X_1(\bar{\boldsymbol{y}})$, hence $\Theta = X_2(\bar{\boldsymbol{y}})$.

Define $N_2(\boldsymbol{y})$ to be the number of elements in $X_2(\boldsymbol{y})$ up to equivalence and define $N_3(\boldsymbol{y})$ and $N_4(\boldsymbol{y})$ to be the number of elements in $X_3(\boldsymbol{y})$ and $X_4(\boldsymbol{y})$ respectively. As is reasonable to expect, Lemma 6 proves that for most $\boldsymbol{y}$, $N_3(\boldsymbol{y}) = N_4(\boldsymbol{y})$. Moreover, provided none of the channel estimates $\hat{\tilde{\boldsymbol{h}}} \in X_2(\boldsymbol{y})$ have their first element equal to zero, $N_2(\boldsymbol{y}) = N_4(\boldsymbol{h})$. Theorem 7 proves that this is almost always the case; the number of least-squares channel estimates up to equivalence almost surely equals the number of pre-images of $\boldsymbol{y}$ under $F$.

**Lemma 6.** *Define $F(\boldsymbol{s}, \boldsymbol{h})$, $X_3(\boldsymbol{y})$, $N_3(\boldsymbol{y})$ and $N_4(\boldsymbol{y})$ as above. Define $N$ to be the generic number of pre-images of $F$. There exists a polynomial $h$ which is not identically zero on the image of $F$ and such that, for any $\boldsymbol{y}$ in the closure of the image of $F$, $h(\boldsymbol{y}) \neq 0$ implies that $N_3(\boldsymbol{y}) = N_4(\boldsymbol{h}) = N$. Furthermore, if $F$ has full rank then it suffices to choose $h$ as in Theorem 1, in which case both the $\boldsymbol{s}$-coordinates and the $\boldsymbol{h}$-coordinates of the points in $X_3(\boldsymbol{y})$ are distinct if $h(\boldsymbol{y}) \neq 0$.*

**Proof.** Assume first that $F$ has full rank and choose $h$ as in Theorem 1, so that $h(\boldsymbol{y}) \neq 0$ implies $N_3(\boldsymbol{y}) = N$. Assume to the contrary that there exist distinct points $(\boldsymbol{s}_1, \boldsymbol{h}_1), (\boldsymbol{s}_2, \boldsymbol{h}_2) \in X_3(\boldsymbol{y})$ with either $\boldsymbol{s}_1 = \boldsymbol{s}_2$ or $\boldsymbol{h}_1 = \boldsymbol{h}_2$. Due to the bi-affine structure of $F$, if $\boldsymbol{h}_1 = \boldsymbol{h}_2$ then $F(\boldsymbol{s}_1, \boldsymbol{h}_1) = F(\boldsymbol{s}_1 + \lambda(\boldsymbol{s}_2 - \boldsymbol{s}_1), \boldsymbol{h}_1)$ for all $\lambda \in \mathbb{C}$. Similarly, if $\boldsymbol{s}_1 = \boldsymbol{s}_2$ then $F(\boldsymbol{s}_1, \boldsymbol{h}_1) = F(\boldsymbol{s}_1, \boldsymbol{h}_1 + \lambda(\boldsymbol{h}_2 - \boldsymbol{h}_1))$ for all $\lambda \in \mathbb{C}$. Both cases contradict $N_3(\boldsymbol{y}) = N < \infty$, proving that the $\boldsymbol{s}$ and $\boldsymbol{h}$ coordinates are distinct, in turn proving that $N_3(\boldsymbol{y}) = N_4(\boldsymbol{y})$.

Assume now that $F$ does not have full rank. Define the variety $W = \{(\boldsymbol{s}, \boldsymbol{h})$: rank $\{HP\} < p\}$ and note that dim $W < p+l$ because it is assumed in Section 2 that $P$ has full column rank. Define $h$ as in Corollary 16 (Appendix A) so that $h(\boldsymbol{y}) \neq 0$ implies that there are an infinite number of pre-images of $\boldsymbol{y}$ lying outside $W$. Let $(\boldsymbol{s}_1, \boldsymbol{h}_1)$ and $(\boldsymbol{s}_2, \boldsymbol{h}_2)$ be two such pre-images, that is, $F(\boldsymbol{s}_1, \boldsymbol{h}_1) = H_1 P \boldsymbol{s}_1 = H_2 P \boldsymbol{s}_2 = F(\boldsymbol{s}_2, \boldsymbol{h}_2)$ with both $H_1 P$ and $H_2 P$ having full column rank. Clearly, if $\boldsymbol{h}_1 = \boldsymbol{h}_2$ then $\boldsymbol{s}_1 = \boldsymbol{s}_2$. Thus $N_4(\boldsymbol{y}) = \infty = N$. □

**Theorem 7.** *As above, let $N(\boldsymbol{s}, \tilde{\boldsymbol{h}}, \boldsymbol{w})$ denote the number of least-squares estimates of the channel up to equivalence. Define $N$ to be the generic number of pre-images of the polynomial map $F$ defined in* (5)*. If there is no noise ($\boldsymbol{w} = 0$) then, for generic $\boldsymbol{s}$ and $\tilde{\boldsymbol{h}}$, $N(\boldsymbol{s}, \tilde{\boldsymbol{h}}, 0) = N$. If additive noise is present then, for $\boldsymbol{w}$ chosen at random and for arbitrary $\boldsymbol{s}$ and $\tilde{\boldsymbol{h}}$, $N(\boldsymbol{s}, \tilde{\boldsymbol{h}}, \boldsymbol{w}) = N$ with probability one.*

**Proof.** The proof continues on from the discussion preceeding Lemma 6. It is proved below that there exists a polynomial $h$ which is not identically zero on the image of $\tilde{F}$ such that, for any $\boldsymbol{y}$ in the closure of the image of $\tilde{F}$, $h(\boldsymbol{y}) \neq 0$ implies $N_2(\boldsymbol{y}) = N$. This proves the theorem in the noise free case because $N(\boldsymbol{s}, \tilde{\boldsymbol{h}}, 0) = N_2(\tilde{F}(\boldsymbol{s}, \tilde{\boldsymbol{h}}))$, and moreover, Lemma 2 implies that, for generic $\boldsymbol{s}$ and $\tilde{\boldsymbol{h}}$, $h(\tilde{F}(\boldsymbol{s}, \tilde{\boldsymbol{h}})) \neq 0$.It also proves the theorem when noise is present because, with probability one, $N(\boldsymbol{s}, \boldsymbol{h}, \boldsymbol{w}) = N_2(\bar{\boldsymbol{y}})$ where $\bar{\boldsymbol{y}}$ is defined in Proposition 5, and moreover, Proposition 5 proves that $h(\bar{\boldsymbol{y}}) \neq 0$ with probability one.

Consider first the case $N = \infty$. Define $W$ to be the set of all points $(\boldsymbol{s}, \tilde{\boldsymbol{h}})$ with the first element of $\tilde{\boldsymbol{h}}$ zero. Corollary 16 applied to $\tilde{F}$ proves that there exists a polynomial $h_1$ such that $h(\boldsymbol{y}) \neq 0$ implies that $\boldsymbol{y}$ is contained in the image of $F$. Lemma 6 proves that there exists an $h_2$ such that, provided $\boldsymbol{y}$ is in the image of $F$, $h_2(\boldsymbol{y}) \neq 0$ implies $N_4(\boldsymbol{y}) = \infty$. Define $h(\boldsymbol{y}) = h_1(\boldsymbol{y}) h_2(\boldsymbol{y})$. Then $h(\boldsymbol{y}) \neq 0$ implies $N_4(\boldsymbol{y}) = \infty$ which in turn implies $N_2(\boldsymbol{y}) = \infty$ because $N_2(\boldsymbol{y}) \geqslant N_4(\boldsymbol{y})$.

The case of $N$ finite is similar but requires the stronger condition in Corollary 16. It is proved below that dim $\overline{\tilde{F}(W)} <$ dim $\overline{\tilde{F}(\mathbb{C}^{p+l+1})}$ if $N$ is finite. (Note that the $N$ in Corollary 16 is still infinite because $\tilde{F}$ cannot have full rank due to the scale ambiguity.) Thus, since $N_2(\boldsymbol{y}) = N_4(\boldsymbol{y})$ if no element of $X_1(\boldsymbol{y})$ lies in $W$, Corollary 16 shows that there exists a polynomial $h_1$ such that $h_1(\boldsymbol{y}) \neq 0$ implies $N_2(\boldsymbol{y}) = N_4(\boldsymbol{y})$. Define $h_2$ as in Lemma 6 so that $h_2(\boldsymbol{y}) \neq 0$ implies $N_4(\boldsymbol{y}) = N$. Thus, $h(\boldsymbol{y}) = h_1(\boldsymbol{y}) h_2(\boldsymbol{y}) \neq 0$ implies $N_2(\boldsymbol{y}) = N$, as required.

To prove dim $\overline{\tilde{F}(W)} <$ dim $\overline{\tilde{F}(\mathbb{C}^{p+l+1})}$ if $N$ is finite, define the sets

$$U_1 = \{\boldsymbol{y}: \boldsymbol{y} = \tilde{F}(\boldsymbol{s}, [0\ 0\ \boldsymbol{h}^{\mathrm{T}}]^{\mathrm{T}}),\ \boldsymbol{s} \in \mathbb{C}^p, \boldsymbol{h} \in \mathbb{C}^{l-1}\}, \tag{8}$$

$$U_2 = \{\boldsymbol{y}: \boldsymbol{y} = \tilde{F}(\boldsymbol{s}, [0\ 1\ \boldsymbol{h}^{\mathrm{T}}]^{\mathrm{T}}),\ \boldsymbol{s} \in \mathbb{C}^p, \boldsymbol{h} \in \mathbb{C}^{l-1}\}, \tag{9}$$

$$U = \{\boldsymbol{y}: \boldsymbol{y} = \tilde{F}(\boldsymbol{s}, [0\ \boldsymbol{h}^{\mathrm{T}}]^{\mathrm{T}}),\ \boldsymbol{s} \in \mathbb{C}^p, \boldsymbol{h} \in \mathbb{C}^{l}\}, \tag{10}$$

$$V = \{\boldsymbol{y}: \boldsymbol{y} = \tilde{F}(\boldsymbol{s}, [1\ \boldsymbol{h}^{\mathrm{T}}]^{\mathrm{T}}),\ \boldsymbol{s} \in \mathbb{C}^p, \boldsymbol{h} \in \mathbb{C}^{l}\}. \tag{11}$$

Since $V$ is the image of the full rank map $F$, Lemma 15 implies that dim $\bar{V} = p+l$. Lemma 15 also implies that dim $\bar{U}_1 < p+l$ and dim $\bar{U}_2 < p+l$. Since $U = U_1 \cup U_2$, dim $\bar{U} < p+l$ too. In particular, dim $\bar{U} <$ dim $\bar{V}$. Since $\tilde{F}(\mathbb{C}^{p+l+1}) = U \cup V$, $\overline{\tilde{F}(\mathbb{C}^{p+l+1})} = \bar{U} \cup \bar{V} = \bar{V}$, the last equality a consequence of $\overline{\tilde{F}(\mathbb{C}^{p+l+1})}$ being an irreducible variety (Lemma 15) and dim $\bar{U} <$ dim $\bar{V}$. Thus dim $\overline{\tilde{F}(W)} =$ dim $\bar{U} <$ dim $\bar{V} =$ dim $\overline{\tilde{F}(\mathbb{C}^{p+l+1})}$, as required. □

**Remark.** Theorem 7 considered the noise free case as well as the additive noise case because, on its own, the statement that $N(\boldsymbol{s}, \boldsymbol{h}, \boldsymbol{w}) = N$ with probability one for arbitrary $\boldsymbol{s}$ and $\boldsymbol{h}$ hides the need for a persistence of excitation condition on $\boldsymbol{s}$ and a regularity condition on $\boldsymbol{h}$ if the resulting channel estimate is to be meaningful. Specifically, assume that the precoder is such that $N = 1$ in Theorem 7. If $\boldsymbol{s} = 0$ and/or $\boldsymbol{h} = 0$ then the channel output is just noise: $\boldsymbol{y} = \boldsymbol{w}$. Even though (4) will almost surely have a unique minimum up to scale if $\boldsymbol{w}$ is chosen at random (that is, $N(0, 0, \boldsymbol{w}) = 1$ almost surely), it is clearly not possible to identify the channel in any sensible way. This is reflected in the noise free statement that $N(\boldsymbol{s}, \boldsymbol{h}, 0) = \infty$ if $\boldsymbol{s} = 0$ and/or $\boldsymbol{h} = 0$. The importance of the noise free result is that it shows that the persistence of excitation and regularity conditions are very mild; for generic $\boldsymbol{s}$ and generic $\boldsymbol{h}$, $N(\boldsymbol{s}, \boldsymbol{h}, 0) = 1$ in this example.

The implication of Theorem 7, irrespective of whether or not additive noise is present, is that the following three cases are exhaustive. The channel is

*identifiable* if $F$ in (5) generically has one pre-image, it is *weakly identifiable* if $F$ generically has more than one but a finite number of pre-images, and it is *not identifiable* if $F$ generically has an infinite number of solutions. The physical meaning of these definitions follows from Theorem 7; in the noise free case, under mild conditions on the source symbols and the channel, the receiver can determine the channel, up to an unknown scaling factor, if and only if the channel is identifiable. If additive noise is present, the least-squares estimate will be unique up to an unknown scaling factor with probability one if and only if the channel is identifiable. Similarly, if the channel is weakly identifiable then there is more than one but a finite number of possibilities for the channel up to scale.

**Remark.** Weak identifiability is still a useful property; the finite number of possible channel estimates might be reduced to a single one by exploiting extra knowledge gained from a finite alphabet constraint on the elements of $\boldsymbol{s}$ or, in an adaptive environment, from an old estimate of the channel $\boldsymbol{h}$. (In the latter case, choose the current estimate to be the one closest to the old estimate.) Note too that since the Jacobian matrix of $F$ is linear in $\boldsymbol{s}$ and $\boldsymbol{h}$, Proposition 3 is a straightforward test for weak identifiability.

## 5. Generic linear precoders

Rather than consider individual precoders, this section considers whole families of precoders and makes statements about almost all members of each family. This enables the big picture to be seen. It is proved that the amount of redundancy that must be introduced to enable the identification of the channel almost always depends on the size of the precoder matrix and not on its individual elements.

The following definitions are introduced for conciseness. For a given channel order $l$ and precoder matrix $P \in \mathbb{C}^{(n+l)\times p}$, define $F(\boldsymbol{s}, \boldsymbol{h}) = HP\boldsymbol{s}$ as in (5). Let $N$ denote the generic number of pre-images of $F$. Then $P$ is said to be *strong* if $N = 1$, it is *weak* if $1 < N < \infty$, and it is *inept* if $N = \infty$. From Theorem 7, it is clear that a strong precoder enables the receiver to identify the channel while a weak precoder enables the receiver to identify weakly the channel.

**Remark.** The strength of a specific precoder can be determined by using symbolic techniques [3] to calculate the generic number of pre-images of $F$.

A property holds for a generic precoder if there exists a non-zero polynomial $g\colon \mathbb{C}^{(n+l)\times p} \to \mathbb{C}$ in the elements of $P$ such that the property holds for all precoders $P$ satisfying $g(P) \neq 0$. If a property holds for a generic precoder then it holds with probability one for a precoder chosen at random.

Proposition 8 and Theorem 9 are the main results of this section and are proved in Section 5.1. Proposition 8 states that the strength of a generic precoder depends only on its size and the channel order. Theorem 9 gives necessary and sufficient conditions for a generic precoder to be strong, weak or inept. The section concludes with two examples showing that non-generic precoders not obeying these rules do exist.

**Proposition 8.** *For any triple* $(n, p, l)$ *there exists a number* $N_{(n,p,l)}$ *such that, for a generic precoder* $P \in \mathbb{C}^{(n+l)\times p}$, *the resulting system* $F$, *defined in* (5), *has* $N_{(n,p,l)}$ *pre-images generically*.

The function $F$ consists of $n$ equations in $p + l$ variables. Excluding the trivial case $p = 1$, it might be anticipated that if $n = p + l$, a generic precoder is strong. Theorem 9 shows that this is not the case; if $n = p + l$ then a generic precoder is weak. Only if the number of equations exceeds the number of variables $(n > p + l)$ is a generic precoder strong. This is due to the special structure of $F$.

**Theorem 9.** *Define* $N_{(n,p,l)}$ *as in Proposition* 8. *Then* $N_{(n,p,l)} = \infty$ *if* $n < p+l$; $1 < N_{(n,p,l)} < \infty$ *if* $n = p+l$ *and* $p > 1$; *or* $N_{(n,p,l)} = 1$ *if* $n > p + l$.

The following exceptions to the rule illustrate the need for considering generic precoders.

**Example 10.** Choose $p$ and $l$ arbitrarily but set $n = p + l$. Consider the precoder $P$ which maps $\boldsymbol{s}$ to $[0, \ldots, 0, s_1, 0, \ldots, 0, s_2, \ldots, s_p]$ where there are $l$ zeros both before and after $s_1$. The elements of the output vector (given by (2) with $h_0 = 1$) satisfy $y_1 = s_1$, $y_2 = h_1 s_1$, $y_3 = h_2 s_1$ and so forth. Therefore, $P$ is strong.

Adding an extra equation to a weak precoder does not necessarily make it strong.

**Example 11.** Set $l=1$, $n=4$ and $p=3$. The precoder which maps $\boldsymbol{s}$ to $[0, s_3, s_1, s_2, s_3]$ can be shown to be weak; the system $F$ generically has 3 pre-images. It might be expected that the precoder of size $n=5$ which maps $\boldsymbol{s}$ to $[0, s_2, s_3, s_1, s_2, s_3]$ is strong now since $n > p+l$. However, the system $F$ still has 3 pre-images generically.

### 5.1. Proofs of Proposition 8 and Theorem 9

If a system $F$ has full rank then adding an extra equation will often make the enlarged system rationally invertible (see Example 11 though). Proposition 12 makes this precise.

**Proposition 12.** *Let $G(\boldsymbol{z}, \boldsymbol{v})$ be a polynomial map decomposable as $G(\boldsymbol{z}, \boldsymbol{v}) = (G_1(\boldsymbol{z}), G_2(\boldsymbol{z}, \boldsymbol{v}))$ where $G_1$ has full rank. If there exists a point $(\bar{\boldsymbol{z}}, \bar{\boldsymbol{v}})$ such that $\bar{\boldsymbol{z}}$ is a generic point of $G_1$ and $G(\bar{\boldsymbol{z}}, \bar{\boldsymbol{v}})$ has a single pre-image then $G$ is rationally invertible.*

**Proof.** A consequence of parts 1 and 2 of Theorem 1 is that it suffices to find open sets $\Omega$ and $\Theta$ such that for all $\boldsymbol{v} \in \Omega$ and $\boldsymbol{z} \in \Theta$, $G(\boldsymbol{z}, \boldsymbol{v})$ has a single pre-image. Let $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N$ be the $N$ pre-images of $G_1(\bar{\boldsymbol{z}})$ with $\boldsymbol{z}_1 = \bar{\boldsymbol{z}}$. For $\varepsilon_1, \varepsilon_2 > 0$ define the sets

$$Y_i = \{G_2(\boldsymbol{z}, \boldsymbol{v}) : \boldsymbol{z} \in B(\boldsymbol{z}_i; \varepsilon_1), \boldsymbol{v} \in B(\bar{\boldsymbol{v}}; \varepsilon_2)\}, \quad i = 1, \ldots, N. \tag{12}$$

The continuity of $G_2$ ensures there exist $\varepsilon_1, \varepsilon_2 > 0$ such that $Y_1 \cap Y_j = \emptyset$ for $j > 1$. Choose $\delta$ as in part 3 of Theorem 1 so that a change in $\bar{\boldsymbol{z}}$ of less than $\delta$ will change each of the pre-images $\boldsymbol{z}_i$ by less than $\varepsilon_1$. Taking $\Omega = B(\bar{\boldsymbol{v}}; \varepsilon_2)$ and $\Theta = B(\bar{\boldsymbol{z}}; \delta)$ completes the proof. □

Define $\boldsymbol{p} \in \mathbb{C}^{(n+l)p}$ to be the vector representation of the precoder matrix $P$, that is, $P = \mathrm{vec}^{-1}\boldsymbol{p}$. The map $F$ in (5) can thus be written as $F(\boldsymbol{s}, \boldsymbol{h}; \boldsymbol{p}) = H(\mathrm{vec}^{-1}\boldsymbol{p})\boldsymbol{s}$. Define the polynomial map $G: \mathbb{C}^{p+l+(n+l)p} \to \mathbb{C}^{n+(n+l)p}$ to be

$$G(\boldsymbol{s}, \boldsymbol{h}, \boldsymbol{p}) = (F(\boldsymbol{s}, \boldsymbol{h}; \boldsymbol{p}), \boldsymbol{p}). \tag{13}$$

**Proof of Proposition 8.** Define $N_{(n,p,l)}$ to be the generic number of pre-images of $G$. Since $G(\boldsymbol{s}', \boldsymbol{h}', \boldsymbol{p}') = G(\boldsymbol{s}, \boldsymbol{h}, \boldsymbol{p})$ is equivalent to $F(\boldsymbol{s}', \boldsymbol{h}') = F(\boldsymbol{s}, \boldsymbol{h})$ and $P' = P$, apply Lemma 2 to conclude that, for generic $P$, $F(\boldsymbol{s}', \boldsymbol{h}') = F(\boldsymbol{s}, \boldsymbol{h})$ has $N_{(n,p,l)}$ solutions for generic $(\boldsymbol{s}, \boldsymbol{h})$. □

**Lemma 13.** *For any $l \geqslant 1$, $p > 1$ and $n = p + l$, there exists a weak precoder $P \in \mathbb{C}^{(n+l)\times p}$.*

**Proof.** Choose $P$ so as to map $\boldsymbol{s}$ to $[0, \ldots, 0, s_1, s_2, 0, \ldots, 0, s_3, \ldots, s_p]$ where there are $l$ zeros both before $s_1$ and between $s_2$ and $s_3$. The elements of the output vector (2) thus satisfy the equations $y_1 = s_1$, $y_2 = s_2 + h_1 s_1$, $y_3 = h_1 s_2 + h_2 s_1$ up to $y_{l+2} = h_l s_2$. These represent $l+2$ equations in $l+2$ unknowns. By Proposition 4, it suffices to find a single output vector for which there is more than one solution. Choose $y_1 = -1$ then. Repeated substitution shows that $s_2$ must satisfy $s_2^{l+1} - y_2 s_2^l - y_3 s_2^{l-1} - \cdots - y_{l+2} = 0$. For generic $(y_2, \ldots, y_{l+2})$ there are $l+1 > 1$ solutions of this equation, and it can be verified that each leads to exactly one solution of the full system $\boldsymbol{y} = F(\boldsymbol{s}, \boldsymbol{h})$. □

**Proof of Theorem 9.** Define $G$ as in (13) and observe from the proof of Proposition 8 that $N_{(n,p,l)}$ is simply the generic number of pre-images of $G$. If $n < p + l$ then $G$ has fewer equations than unknowns and hence an infinite number of pre-images; $N_{(n,p,l)} = \infty$ if $n < p + l$.

Assume $n = p + l$ and $p > 1$. Lemma 13 proves that there exists a point $(\boldsymbol{s}, \boldsymbol{h}, \boldsymbol{p})$ such that $G(\boldsymbol{s}, \boldsymbol{h}, \boldsymbol{p})$ has more than one but less than an infinite number of pre-images. Thus, Proposition 4 implies that $1 < N_{(n,p,l)} < \infty$.

Assume $n > p + l$. The following proof that $G$ is rationally invertible exploits the fact that the last element $x_n$ of the encoded vector $\boldsymbol{x} = P\boldsymbol{s}$ affects only the last element $y_n$ of the output vector $\boldsymbol{y} = H\boldsymbol{x}$. Partition the matrices $H$ and $P$ as follows:

$$HP\boldsymbol{s} = \begin{bmatrix} H_1 & 0 \\ \boldsymbol{u}^{\mathrm{T}} & 1 \end{bmatrix} \begin{bmatrix} P_1 \\ \boldsymbol{v}^{\mathrm{T}} \end{bmatrix} \boldsymbol{s} = \begin{bmatrix} H_1 P_1 \boldsymbol{s} \\ \boldsymbol{u}^{\mathrm{T}} P_1 \boldsymbol{s} + \boldsymbol{v}^{\mathrm{T}} \boldsymbol{s} \end{bmatrix}, \tag{14}$$

where $\boldsymbol{u}^{\mathrm{T}} = [0 \cdots 0 \; h_l \cdots h_1]$ and $\boldsymbol{v}^{\mathrm{T}}$ is the last row of $P$. Let $\boldsymbol{p}_1$ be the vector representation of $P_1$, that is, $P_1 = \mathrm{vec}^{-1}\boldsymbol{p}_1$. Decompose the map $G$ accordingly

$$G(\boldsymbol{s}, \boldsymbol{h}, \boldsymbol{p}) = (G_1(\boldsymbol{s}, \boldsymbol{h}, \boldsymbol{p}_1), G_2(\boldsymbol{s}, \boldsymbol{h}, \boldsymbol{p}_1, \boldsymbol{v})), \tag{15}$$

$$G_1 = (H_1(\mathrm{vec}^{-1}\boldsymbol{p}_1)\boldsymbol{s}, \boldsymbol{p}_1), \tag{16}$$

$$G_2 = ((\boldsymbol{u}^{\mathrm{T}}(\mathrm{vec}^{-1}\boldsymbol{p}_1)\boldsymbol{s} + \boldsymbol{v}^{\mathrm{T}}\boldsymbol{s}), \boldsymbol{v}). \tag{17}$$

Notice that $G_1$ is identical to $G$ in (13) if the precoder $P_1$ were used instead of $P$. Proposition 4 implies that $G_1$ has full rank since for any $n > p + l$ there exists a $P_1 \in \mathbb{C}^{(n+l-1)\times p}$ which is a weak precoder (take for instance the precoder in Example 10 with $n-(p+l)-1$ zeros appended). In order to apply Proposition 12, let $(\bar{\boldsymbol{s}}, \bar{\boldsymbol{h}}, \bar{\boldsymbol{p}}_1)$ be a generic point of $G_1$ and define $(\boldsymbol{s}_i, \boldsymbol{h}_i)$ so that

$$\begin{aligned}&\{(\boldsymbol{s}_1, \boldsymbol{h}_1), \ldots, (\boldsymbol{s}_N, \boldsymbol{h}_N)\}\\ &\quad = \{(\boldsymbol{s}, \boldsymbol{h})\colon G_1(\boldsymbol{s}, \boldsymbol{h}, \bar{\boldsymbol{p}}_1) = G_1(\bar{\boldsymbol{s}}, \bar{\boldsymbol{h}}, \bar{\boldsymbol{p}}_1)\}.\end{aligned} \tag{18}$$

Note that the last sentence of Lemma 6 ensures that $\boldsymbol{s}_i \neq \boldsymbol{s}_j$ for $i \neq j$. For each $(\boldsymbol{s}_i, \boldsymbol{h}_i)$ the first component of $G_2$ takes the value $\boldsymbol{u}_i^{\mathrm{T}} \bar{P}_1 \boldsymbol{s}_i + \boldsymbol{v}^{\mathrm{T}}\boldsymbol{s}_i$ (where $\boldsymbol{u}_i$ depends only on $\boldsymbol{h}_i$). These values can be made distinct by a judicious choice of $\boldsymbol{v}$. By Proposition 12, $G$ is rationally invertible. □

### 5.2. Generic zero prefix precoders

A zero prefix precoder is a precoder $P \in \mathbb{C}^{(n+l)\times p}$ whose first $l$ rows are zero. Such a precoder sets the initial state of the channel to zero, that is, $x_{1-l} = \cdots = x_0 = 0$ in (2). This is arguably a nice thing to do. Since a randomly chosen precoder will not have a zero prefix with probability one, there is no reason for Proposition 8 and Theorem 9 to hold for zero prefix precoders. Moreover, it is plausible that the condition $n > p + l$ in Theorem 9 can be relaxed for zero prefix precoders. However, this is not the case. This is formalised below.

A property holds for a generic zero prefix precoder if there exists a non-zero polynomial $g\colon \mathbb{C}^{(n+l)\times p} \to \mathbb{C}$ such that the property holds for all precoders $P$ satisfying $g(P) \neq 0$ and whose first $l$ rows are zero.

**Theorem 14.** *For any triple $(n, p, l)$ there exists a number $N^{\mathrm{zp}}_{(n,p,l)}$ such that, for a generic zero prefix precoder $P \in \mathbb{C}^{(n+l)\times p}$, the system $F$ defined in* (5) *has $N^{\mathrm{zp}}_{(n,p,l)}$ pre-images generically. Moreover, $N^{\mathrm{zp}}_{(n,p,l)} = \infty$ if $n < p+l$; $1 < N^{\mathrm{zp}}_{(n,p,l)} < \infty$ if $n = p+l$ and $p > 1$; or $N^{\mathrm{zp}}_{(n,p,l)} = 1$ if $n > p + l$.*

**Proof.** By design, only a minor modification of the proofs in Section 5.1 is required. □

**Remark.** Whether or not $N^{\mathrm{zp}}_{(n,p,l)} = N_{(n,p,l)}$ holds for all $(n, p, l)$, where $N_{(n,p,l)}$ is defined in Proposition 8, is not investigated here.

## 6. Conclusion

This paper drew attention to the fact that, in a linearly precoded wireless communication system, the impulse response of channel (1) can often be determined by solving a system of noise-corrupted polynomial equations. In Section 4 it was proved that the number of channel estimates obtained by solving this system of equations in the least-squares sense can be determined by studying the generic number of solutions of a related polynomial equation. Standard results from algebraic geometry were then applied in Section 5 to prove that the feasibility of the receiver identifying the channel is governed primarily by the size of the precoder matrix and not, in general, on the individual elements of the precoder matrix.

## Appendix A. Technical results on polynomial maps

The following known results on polynomial maps are not included in Section 3 because knowledge of these results is not essential on a first reading. These results require the definitions of a variety,[2] an irreducible variety, and the dimension of a variety, all of which can be found in [3].

**Lemma 15.** *Let $F\colon \mathbb{C}^m \to \mathbb{C}^n$ be a polynomial map and let $W \subset \mathbb{C}^m$ be a variety. Then $U = \overline{F(W)}$, the closure of the image of $W$, is also a variety. Moreover, if $W$ is irreducible then so too is $U$. Also,* $\dim U \leqslant \dim W$. *If $W = \mathbb{C}^m$ then* $\dim U = m$ *if and only if $F$ has full rank.*

Corollary 16 is an extension of Theorem 1 and can be deduced from Lemma 15. Note that $N$ is allowed to be infinite.

[2] The word variety is used here to mean an algebraic set and is consistent with the usage in [3]. However, the reader should be aware that this differs from the modern terminology which defines a variety to be an irreducible algebraic set.

**Corollary 16.** *Let* $F\colon \mathbb{C}^m \to \mathbb{C}^n$ *be a polynomial map which generically has* $N$ *pre-images and let* $W \subset \mathbb{C}^m$ *be a variety with* $\dim W < m$. *There exists a polynomial* $h\colon \mathbb{C}^n \to \mathbb{C}$ *which is not identically zero on* $F(\mathbb{C}^m)$ *and such that, for any* $\boldsymbol{y} \in \overline{F(\mathbb{C}^m)}$, $h(\boldsymbol{y}) \neq 0$ *implies that there are* $N$ *pre-images of* $\boldsymbol{y}$ *under* $F$ *that lie outside* $W$. *This can be strengthened if* $\dim \overline{F(W)} < \dim \overline{F(\mathbb{C}^m)}$ (*which is always the case if* $N$ *is finite*), *in which case* $h(\boldsymbol{y}) \neq 0$ *also implies there are no pre-images of* $\boldsymbol{y}$ *under* $F$ *contained in* $W$.

## References


[1] M.A. Armstrong, Basic Topology, Springer, Berlin, 1983.

[2] H. Bass, E.H. Connell, D. Wright, The Jacobian conjecture: reduction of degree and formal expansion of the inverse, Bull. Amer. Math. Soc. 7 (1982) 287–330.

[3] D.A. Cox, J.B. Little, D. O'Shea, Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra, second ed., Springer, Berlin, 1996.

[4] P. Griffiths, J. Harris, Principles of Algebraic Geometry, Wiley, New York, 1978.

[5] F. Gustafsson, B. Wahlberg, Blind equalization by direct examination of the input sequences, IEEE Trans. Comm. 43 (7) (1995) 2213–2222.

[6] J. Harris, Algebraic Geometry: A First Course, Springer, Berlin, 1992.

[7] H. Liu, G. Xu, L. Tong, T. Kailath, Recent developments in blind channel equalization: from cyclostationarity to subspaces, Signal Process. 50 (1996) 83–99.

[8] J.H. Manton, Dissecting OFDM: the independent roles of the cyclic prefix and the IDFT operation, IEEE Comm. Lett. 5 (12) (2001) 474–476.

[9] J.H. Manton, Finite alphabet source recovery in polynomial systems, Systems Control Lett. 47 (4) (2002) 279–286.

[10] J.H. Manton, An improved least-squares blind channel identification algorithm for linearly and affinely precoded communication systems, IEEE Signal Process. Lett. 9 (9) (2002) 282–285.

[11] J.H. Manton, An OFDM interpretation of zero padded block transmissions, Systems Control Lett. 47 (5) (2002) 393–399.

[12] J.H. Manton, Design and analysis of linear precoders under a mean square error criterion, part I: foundations and worst case designs, Systems Control Lett. 49 (2) (2003) 121–130.

[13] J.H. Manton, Design and analysis of linear precoders under a mean square error criterion, part II: MMSE designs and conclusions, Systems Control Lett. 49 (2) (2003) 131–140.

[14] J.H. Manton, J.R.J. Groves, Y. Hua, On properties of the solutions of systems of polynomial equations, in:The Third Asian Control Conference, Shanghai, China, July 2000.

[15] J.H. Manton, Y. Hua, A frequency domain deterministic approach to channel identification, IEEE Signal Process. Lett. 6 (12) (1999) 323–326.

[16] I.R. Shafarevich, Basic Algebraic Geometry:Varieties in Projective Space, vol. 1, second ed., Springer, Berlin, 1994.

[17] E.D. Sontag, On the observability of polynomial systems, I: finite-time problems, SIAM J. Control Optim. 17 (1) (1979) 139–151.

[18] L. Tong, S. Perreau, Multichannel blind identification: from subspace to maximum likelihood methods, Proc. IEEE 86 (10) (1998) 1951–1968.

[19] A. van der Veen, S. Talwar, A. Paulraj, A subspace approach to blind space–time signal processing for wireless communication systems, IEEE Trans. Signal Process. 45 (1) (1997) 173–190.

[20] G. Zhou, X.-G. Xia, Ambiguity resistant polynomial matrices, Linear Algebra Appl. 286 (1999) 19–35.