# An Improved Least Squares Blind Channel Identification Algorithm for Linearly and Affinely Precoded Communication Systems

Jonathan H. Manton, *Member, IEEE*

***Abstract*—Certain linear and affine precoders introduce enough algebraic redundancy to enable the receiver to identify a single-input single-output finite-impulse response channel without making any statistical assumptions on the source sequence. However, quite surprisingly, the traditional steepest descent least squares algorithm for estimating the channel often fails to converge, even in the absence of noise. This letter explains why this is the case and derives a novel steepest descent algorithm on complex projective space that is guaranteed to converge. The complex projective space formulation also provides a standard framework for understanding different performance measures proposed in the literature.**

***Index Terms*—Algebraic channel identification, bilinear equations, linear precoders, optimization on manifolds, wireless communications.**

## I. Introduction

Certain linear and affine precoders [5], [9] allow the receiver to identify blindly a single-input single-output finite-impulse response (SISO-FIR) channel based only on the algebraic redundancy introduced by the precoder [4] and, in particular, without any statistical assumptions on the transmitted symbols. Specifically, in the presence of additive noise, the channel can be estimated blindly by solving a nonlinear least squares problem. A standard approach for solving nonlinear least squares problems is to use a steepest descent algorithm. It is known that under very mild conditions, a steepest descent algorithm coupled with Armijo's step size rule is guaranteed to converge to a critical point [8]. It is, thus, surprising that it often fails to converge (at least in a reasonable number of iterations) when applied to the channel estimation problem. This letter explains why this is the case and derives a novel steepest descent algorithm that is guaranteed to converge. In doing so, it also provides a unified framework in which to understand the different performance measures proposed in the literature [7], [10].

The signal model is as follows. A vector of $p$ complex-valued source symbols $\boldsymbol{s} \in \mathbb{C}^p$ is affinely precoded [5] to form $n+L-1$ complex-valued encoded symbols $\boldsymbol{x} \in \mathbb{C}^{n+L-1}$ according to the rule

$$\boldsymbol{x} = A\boldsymbol{s} + \boldsymbol{b} \quad A \in \mathbb{C}^{(n+L-1)\times p} \quad \boldsymbol{b} \in \mathbb{C}^{n+L-1} \tag{1}$$

where $A$ and $\boldsymbol{b}$ are known to the receiver, and $L$ is an upper bound on the length of the SISO-FIR channel through which $\boldsymbol{x}$ is transmitted. Different choices of the matrix $A$ allow (1) to model many popular precoding techniques, including filter bank precoders and orthogonal frequency division multiplexing (OFDM) schemes. The vector $\boldsymbol{b}$ allows for the inclusion of a training sequence or pilot tones to assist in channel identification. (When applied to block transmission systems, the vector $\boldsymbol{s}$ is taken to be one or more blocks of source symbols depending on how many blocks are to be used in the identification of the channel.)

In the noise-free case, the $i$th element of the received output vector $\boldsymbol{y} \in \mathbb{C}^n$ is the convolution of the encoded symbols $\boldsymbol{x}$ with the unknown channel vector $\boldsymbol{h} = [h_0, \ldots, h_{L-1}]^T \in \mathbb{C}^L$, namely

$$\boldsymbol{y}_i = \sum_{k=0}^{L-1} h_k \boldsymbol{x}_{i-k+L-1}. \tag{2}$$

By introducing an $n$-by-$(n+L-1)$ upper triangular Toeplitz matrix $H$ having $[h_{L-1}, h_{L-2}, \ldots, h_0, 0, \ldots, 0]$ as its first row, (2) can be written in matrix form as

$$\boldsymbol{y} = H\boldsymbol{x} = H(A\boldsymbol{s} + \boldsymbol{b}). \tag{3}$$

In practice, the received signal $\boldsymbol{y}$ is corrupted by additive noise. It is, therefore, appropriate to solve (3) for both $\boldsymbol{s}$ and $\boldsymbol{h}$ in the least squares sense. Specifically, this letter studies the problem of minimizing the nonlinear least squares cost function $\psi : \mathbb{C}^L \to \mathbb{R}$ defined to be

$$\psi(\boldsymbol{h}) = \min_{\boldsymbol{s}\in\mathbb{C}^p} \phi(\boldsymbol{h}, \boldsymbol{s}) \quad \phi(\boldsymbol{h}, \boldsymbol{s}) = \frac{1}{2}\|\boldsymbol{y} - H(A\boldsymbol{s} + \boldsymbol{b})\|^2. \tag{4}$$

It is remarked that (2) is a bilinear equation, and in particular, the results of this letter apply to the wider problem of solving general bilinear equations in the least squares sense. It is also remarked that conditions for (4) to have a unique minimum are discussed in [6].

The motivation for considering this problem is that often the structure (1), present in many communication systems, is not fully exploited for the purposes of channel identification. A natural question, then, is whether or not the least squares channel

Manuscript received November 3, 2000; revised May 10, 2002. This work was supported by the Australian Research Council. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Marcelo G. S. Bruno.

The author is with the Australian Research Council Special Research Centre for Ultra-Broadband Information Networks, Department of Electrical and Electronic Engineering, The University of Melbourne, Victoria 3010, Australia (e-mail: j.manton@ee.mu.oz.au).

Digital Object Identifier 10.1109/LSP.2002.803613.

estimate obtained by minimizing $\psi(\boldsymbol{h})$ is to be preferred over more traditional channel estimates. Before this question can be answered though, it is necessary for an algorithm to be developed that minimizes $\psi(\boldsymbol{h})$. Surprisingly, the author discovered the standard steepest descent method fails to minimize $\psi(\boldsymbol{h})$. The main contribution of this letter, besides an explanation of why the standard steepest descent algorithm fails to converge, is the derivation of a novel steepest descent algorithm on complex projective space for minimizing $\psi(\boldsymbol{h})$.

It is candidly stated that the advantages and disadvantages of using the proposed least squares channel estimate are not investigated in this letter but rather are a topic of future research. Indeed, the merits of a least squares channel estimator can only be judged with respect to a specific affine precoder, whereas the focus of this letter is on arbitrary affine precoders.

This section concludes with a brief discussion on the scale ambiguity problem associated with all linear and some affine precoders. If a linear precoder is used, then it is not possible to determine the magnitude of the channel vector because, for any nonzero $\lambda \in \mathbb{C}$, $\boldsymbol{y} = \boldsymbol{h} * A\boldsymbol{s} = (\lambda \boldsymbol{h}) * A(\lambda^{-1}\boldsymbol{s})$, where $*$ denotes the convolution operation in (2). (Recall that both $\boldsymbol{h}$ and $\boldsymbol{s}$ are unknown to the receiver.) Moreover, the affine precoder (1) removes the scale ambiguity if and only if $\boldsymbol{b}$ is orthogonal to the range space of $A$, i.e., $(I - AA^{+})\boldsymbol{b} \neq 0$, where $A^{+}$ is the Moore–Penrose pseudoinverse of $A$; thus, $AA^{+}$ denotes projection onto the range space of $A$. The proposed algorithm handles both cases: it identifies the channel up to an unknown scaling factor if $(I - AA^{+})\boldsymbol{b} = 0$, and it identifies the channel completely if $(I - AA^{+})\boldsymbol{b} \neq 0$.

## II. Standard Least Squares

This section derives a standard steepest descent algorithm for estimating the channel vector under the assumption that there is no scale ambiguity. Then, it is explained why this algorithm often fails to converge in a reasonable number of iterations. The results of this section are used in Section III to derive a novel algorithm for channel estimation.

Referring to (4), it is well known that for a fixed $\boldsymbol{h}$, $\phi(\boldsymbol{h}, \boldsymbol{s})$ achieves its minimum when $\boldsymbol{s}$ equals

$$\widehat{\boldsymbol{s}} = \left(A^H H^H H A\right)^{-1} A^H H^H (\boldsymbol{y} - H\boldsymbol{b}). \tag{5}$$

Therefore, the aim is to find the channel $\boldsymbol{h}$ which minimizes $\psi(\boldsymbol{h}) = \phi(\boldsymbol{h}, \widehat{\boldsymbol{s}})$. To apply the steepest descent method, it is necessary to calculate the derivative of $\psi$ with respect to $\boldsymbol{h}$.

The notation $\phi_{\boldsymbol{h}}(\boldsymbol{h}) = (\partial\phi/\partial\Re\boldsymbol{h}) + \jmath(\partial\phi/\partial\Im\boldsymbol{h})$ is used to represent the derivative of $\phi$ with respect to $\boldsymbol{h}$, and similarly for $\psi$. Since $\phi_{\boldsymbol{s}}(\boldsymbol{h}, \widehat{\boldsymbol{s}}) = 0$, the derivative of $\psi(\boldsymbol{h}) = \phi(\boldsymbol{h}, \widehat{\boldsymbol{s}})$ is $\psi_{\boldsymbol{h}}(\boldsymbol{h}) = \phi_{\boldsymbol{h}}(\boldsymbol{h}, \widehat{\boldsymbol{s}})$. Explicitly, if $J_i = \partial H/\partial h_i$ is the $n$-by-$(n + L - 1)$ matrix of the same form as $H$ but with all elements zero except for the $h_i$ terms which are set to one, then

$$\psi_{\boldsymbol{h}}(\boldsymbol{h}) = -\begin{bmatrix} (A\widehat{\boldsymbol{s}} + \boldsymbol{b})^H J_0^T (\boldsymbol{y} - H(A\widehat{\boldsymbol{s}} + \boldsymbol{b})) \\ \vdots \\ (A\widehat{\boldsymbol{s}} + \boldsymbol{b})^H J_{L-1}^T (\boldsymbol{y} - H(A\widehat{\boldsymbol{s}} + \boldsymbol{b})) \end{bmatrix}. \tag{6}$$

Based on (6), a steepest descent algorithm using Armijo's step size rule (see [8, Sec. 1.2.3]) is readily implementable. Armijo's step size rule ensures that the cost function monotonically decreases and, under mild conditions (the cost function is continuously differentiable, bounded from below, and has compact sublevel sets), the algorithm converges to a critical point (a point where $\psi_{\boldsymbol{h}}(\boldsymbol{h}) = 0$). Surprisingly, then, simulations using randomly generated affine precoders showed that the steepest descent algorithm often failed to converge (the gradient $\psi_{\boldsymbol{h}}(\boldsymbol{h})$ did not approach zero). When the algorithm failed to converge, it was observed that the norm of $\boldsymbol{h}$ decreased at each iteration.

The explanation is that $\psi(\boldsymbol{h})$ is discontinous at the origin and, thus, violates the conditions for guaranteed convergence. Moreover, it is not uncommon that, for a given $\boldsymbol{h}$, $\psi(\lambda\boldsymbol{h})$ decreases as $\lambda$ approaches zero from above; roughly speaking, this occurs when $-\boldsymbol{h}$ is a better estimate than $\boldsymbol{h}$ is, for instance. Thus, the steepest descent algorithm can get pulled, increasingly slowly, toward the origin. (It may eventually escape, but convergence is intolerably slow even if it does.)

## III. Least Squares in Complex Projective Space

Section II showed that the traditional steepest descent algorithm often fails to converge because $\psi(\boldsymbol{h})$ is discontinuous at the origin. This section remedies this by minimizing $\psi(\boldsymbol{h})$ on complex projective space rather than on $\mathbb{C}^L$. Furthermore, it is shown that when scale ambiguity is present, complex projective space is the natural framework in which to study the channel estimation problem.

Complex projective space of dimension $L - 1$, denoted $\mathbb{CP}^{L-1}$, is the collection of equivalence classes of $\mathbb{C}^L - \{0\}$, where $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{C}^L - \{0\}$ are equivalent if there exists a $\lambda \in \mathbb{C}$ such that $\boldsymbol{v} = \lambda\boldsymbol{w}$. An element of $\mathbb{CP}^{L-1}$ is written as $\lfloor \boldsymbol{v} \rfloor$ where $\boldsymbol{v} \in \mathbb{C}^L - \{0\}$. Complex projective space is fundamental to algebraic geometry.

Consider first the case when $(I - AA^{+})\boldsymbol{b} \neq 0$, i.e., there is no scale ambiguity. Define $f : \mathbb{C}^L - \{0\} \to \mathbb{R}$ to be $f(\boldsymbol{h}) = \min_{\lambda\in\mathbb{C}} \psi(\lambda\boldsymbol{h})$. Substituting (5) into (4) shows that

$$\begin{aligned} \psi(\lambda\boldsymbol{h}) &= \frac{1}{2}\|Q_{\boldsymbol{h}}\boldsymbol{y} - \lambda Q_{\boldsymbol{h}} H\boldsymbol{b}\|^2 \\ Q_{\boldsymbol{h}} &= I - HA\left(A^H H^H H A\right)^{-1} A^H H^H \end{aligned} \tag{7}$$

where $I$ is the identity matrix. Thus, $\psi(\lambda\boldsymbol{h})$ achieves its minimum when $\lambda$ equals

$$\widehat{\lambda} = \left(\boldsymbol{b}^H H^H Q_{\boldsymbol{h}} H\boldsymbol{b}\right)^{-1} \boldsymbol{b}^H H^H Q_{\boldsymbol{h}}\boldsymbol{y}. \tag{8}$$

It can be shown that since $(I - AA^{+})\boldsymbol{b} \neq 0$, $\boldsymbol{b}^H H^H Q_{\boldsymbol{h}} H\boldsymbol{b}$ is nonzero for all $\boldsymbol{h} \in \mathbb{C}^L - \{0\}$, proving that $f$ is a smooth function of $\boldsymbol{h}$.

The discontinuity of $\psi$ at the origin is removed by considering instead the cost function $\tilde{f} : \mathbb{CP}^{L-1} \to \mathbb{R}$ defined by $\tilde{f}(\lfloor \boldsymbol{h} \rfloor) = f(\boldsymbol{h})$. (It is well defined because $f(\lambda\boldsymbol{h}) = f(\boldsymbol{h})$ for nonzero $\lambda$.) A steepest descent algorithm for minimizing the smooth cost function $\tilde{f}(\lfloor \boldsymbol{h} \rfloor)$ on the compact manifold $\mathbb{CP}^{L-1}$ is now derived. It is guaranteed to converge to a critical point because the conditions for Armijo's step size rule to converge are now satisfied (in particular, the cost function is no longer discontinuous).

Analogously to Section II, the derivative of $f(\boldsymbol{h})$ can be shown to be

$$f_{\boldsymbol{h}}(\boldsymbol{h}) = \widehat{\lambda}^H \psi_{\boldsymbol{h}}(\widehat{\lambda}\boldsymbol{h}) \tag{9}$$

where $\widehat{\lambda}^H$ is the complex conjugate of $\widehat{\lambda}$, and $\widehat{\lambda}$ is defined in (8). The tangent space $T_{\lfloor \boldsymbol{h} \rfloor}$ at $\lfloor \boldsymbol{h} \rfloor \in \mathbb{CP}^{L-1}$ is the $(L-1)$-dimensional [$(L-1)$-D] subspace $T_{\lfloor \boldsymbol{h} \rfloor} = \left\{ \boldsymbol{z} : \boldsymbol{h}^H \boldsymbol{z} = 0 \right\}$ of $\mathbb{C}^L$. The gradient [1] of $\tilde{f}$ at a point $\lfloor \boldsymbol{h} \rfloor$ is only defined once $T_{\lfloor \boldsymbol{h} \rfloor}$ is given an inner product structure. Choosing the standard (Euclidean) inner product for $T_{\lfloor \boldsymbol{h} \rfloor}$ results in the gradient being $(I - \boldsymbol{h}\boldsymbol{h}^H) f_{\boldsymbol{h}}(\boldsymbol{h})$, which is the projection of the gradient of $f$ onto the tangent space $T_{\lfloor \boldsymbol{h} \rfloor}$. (See [2] and [3] for the theory behind optimization on manifolds.) This leads to the following algorithm for finding an $\boldsymbol{h}$ that (locally) minimizes $f(\boldsymbol{h})$. The actual least squares channel estimate $\widehat{\boldsymbol{h}}$ minimizing $\psi(\boldsymbol{h})$ is, thus, $\widehat{\boldsymbol{h}} = \widehat{\lambda}\boldsymbol{h}$ where $\widehat{\lambda}$ is defined in (8).

**Algorithm 1**

1. *Choose* $\boldsymbol{h} \in \mathbb{C}^L$ *such that* $\|\boldsymbol{h}\| = 1$. *Set the step size* $\gamma := 1$.
2. *Compute the steepest descent direction* $z = -\left(I - \boldsymbol{h}\boldsymbol{h}^H\right) f_{\boldsymbol{h}}(\boldsymbol{h})$. *If* $\|z\|$ *is sufficiently small then stop.*
3. *If* $f(\boldsymbol{h}) - f(\boldsymbol{h} + 2\gamma z) \geq \gamma \|z\|^2$ *then set* $\gamma := 2\gamma$ *and repeat Step 3.*
4. *If* $f(\boldsymbol{h}) - f(\boldsymbol{h} + \gamma z) < (1/2)\gamma \|z\|^2$ *then set* $\gamma := (1/2)\gamma$ *and repeat Step 4.*
5. *Set* $\boldsymbol{h} := (\boldsymbol{h} + \gamma z)/(\|\boldsymbol{h} + \gamma z\|)$. *Go to Step 2.*

*Remarks:* Steps 3 and 4 in Algorithm 1 implement the Armijo step size rule [8]. Since $\tilde{f}$ is differentiable and $\mathbb{CP}^{L-1}$ is compact, Algorithm 1 is guaranteed to converge monotonically to a critical point of the least squares cost function [8].

The case when $(I - AA^+)\boldsymbol{b} = 0$ is now considered. Since $(I - AA^+)\boldsymbol{b} = 0$ implies $\psi(\lambda\boldsymbol{h}) = \psi(\boldsymbol{h})$ for any nonzero $\lambda$, the cost function $\tilde{f} : \mathbb{CP}^{L-1} \to \mathbb{R}$ simplifies to $\tilde{f}(\lfloor \boldsymbol{h} \rfloor) = \psi(\boldsymbol{h})$. Thus, Algorithm 1 can be used to estimate the channel by replacing $f$ with $\psi$ and $f_{\boldsymbol{h}}(\boldsymbol{h})$ with $\psi_{\boldsymbol{h}}(\boldsymbol{h})$.

## IV. Performance Measures

It is the secondary purpose of this letter to draw attention to the fact that the natural setting for channel estimation in the presence of scale ambiguity is in complex projective space. That is, rather than consider the channel vector $\boldsymbol{h}$ as an element of $\mathbb{C}^L$, it is best considered as representing the equivalence class $\lfloor \boldsymbol{h} \rfloor$ of all channel vectors agreeing with $\boldsymbol{h}$ up to scale.

The problem of measuring the accuracy of a channel estimate when there is inherent scale ambiguity has been addressed in [7] and [10], where a number of different definitions of distance between the true and estimated channel vectors were proposed. It is interesting to note the equivalence of these distances and the distance functions (the more familiar term "metric" in other branches of mathematics has a different meaning in differential geometry [1]) commonly used in complex projective space.

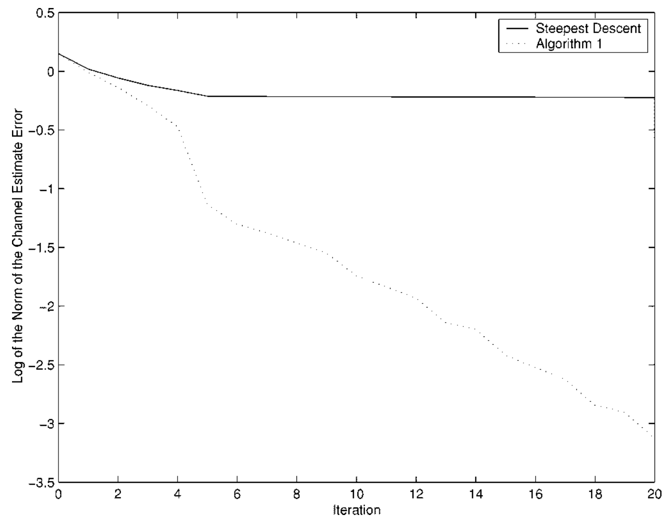


Fig. 1. Graph showing the superior performance of Algorithm 1 over the steepest descent algorithm (Section II).

(In other words, the performance measures used in [7] and [10] are natural measures to use from a mathematical perspective because they are true distance functions on complex projective space.) Moreover, the differences in the distances defined in [7] and [10] potentially can be better understood by studying the induced distance functions on $\mathbb{CP}^{L-1}$. (For instance, some of the distance functions on $\mathbb{CP}^{L-1}$ appearing in [2, Sec. 4.3] are asymptotically equivalent.)

*Remark:* An example of a distance function on $\mathbb{CP}^{L-1}$ is the Fubini–Study distance defined as follows. Let $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{C}^L$ with $\|\boldsymbol{v}\| = \|\boldsymbol{w}\| = 1$ represent the points $\lfloor \boldsymbol{v} \rfloor, \lfloor \boldsymbol{w} \rfloor$ in $(L-1)$-D projective space. The Fubini–Study distance between $\lfloor \boldsymbol{v} \rfloor$ and $\lfloor \boldsymbol{w} \rfloor$ is

$$d\big(\lfloor \boldsymbol{v} \rfloor, \lfloor \boldsymbol{w} \rfloor\big) = \cos^{-1} |\boldsymbol{v}^H \boldsymbol{w}|. \tag{10}$$

Note that since $\min_{\theta \in \mathbb{R}} \|\boldsymbol{v} - e^{\jmath\theta}\boldsymbol{w}\|^2 = 2\big(1 - |\boldsymbol{v}^H\boldsymbol{w}|\big)$, $|\boldsymbol{v}^H\boldsymbol{w}|$ provides a sensible indication of how far apart $\lfloor \boldsymbol{v} \rfloor$ and $\lfloor \boldsymbol{w} \rfloor$ are.

## V. Numerical Example

This section presents a simple example illustrating that the traditional steepest descent algorithm can fail to converge in a reasonable number of iterations. When applied to the same example, Algorithm 1 converges to the correct answer.

The following affine precoder [5] was used (in Matlab, $A = [\text{zeros}(5, 24);\ \text{kron}(\text{eye}(6), \text{eye}(5, 4));\ \text{zeros}(2, 24)]$):

$$A = \begin{bmatrix} 0_{5\times 24} \\ I_6 \otimes I_{5\times 4} \\ 0_{2\times 24} \end{bmatrix} \quad \boldsymbol{b} = \begin{bmatrix} 0_{2\times 1} \\ 1 \\ 0_{34\times 1} \end{bmatrix}. \tag{11}$$

The true channel was chosen to be $\boldsymbol{h} = [1, 1, 1]^T$. Fig. 1 compares the convergence rates of the standard steepest descent algorithm and Algorithm 1. The true source symbols were $\boldsymbol{s} = [1, \ldots, 24]^T$, and the initial channel estimate was $\widehat{\boldsymbol{h}} = [1, 0, 0]^T$. Other examples, not presented here, exhibited similar behavior.

*Remark:* It is interesting to note that for this particular example, the steepest descent algorithm eventually converged to the correct answer after a few thousand iterations.

## VI. Conclusion

Section II explained why the traditional steepest descent least squares algorithm often fails to converge. Section III remedied this lack of convergence by introducing scale ambiguity into the cost function and then minimizing the new cost function on complex projective space. Section IV drew attention to the fact that complex projective space is the natural framework to use in the presence of scale ambiguity. Finally, a numerical example in Section V supported the claims made in this letter.

## References

[1] W. M. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry*, 2nd ed. Orlando, FL: Academic, 1986.

[2] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 2, pp. 303–353, 1998.

[3] J. H. Manton, "Optimization algorithms exploiting unitary constraints," *IEEE Trans. Signal Processing*, vol. 50, pp. 635–650, Mar. 2002.

[4] J. H. Manton and Y. Hua, "Blind channel identifiability with an arbitrary linear precoder," in *Signal Processing Advances in Wireless and Mobile Communications*. Englewood Cliffs, NJ: Prentice-Hall, 2000, vol. 1, ch. 10, pp. 339–366.

[5] J. H. Manton, I. M. Y. Mareels, and Y. Hua, "Affine precoders for reliable communications," in *Proc. ICASSP*, vol. V, Istanbul, Turkey, June 2000, pp. 2749–2752.

[6] J. H. Manton, W. D. Neumann, and P. T. Norbury, "On the algebraic identifiability of finite impulse response channels driven by linearly precoded signals," *Syst. Control Lett.*, 2002, submitted for publication.

[7] D. R. Morgan, J. Benesty, and M. M. Sondhi, "On the evaluation of estimated impulse responses," *IEEE Signal Processing Lett.*, vol. 5, pp. 174–176, July 1998.

[8] E. Polak, *Optimization: Algorithms and Consistent Approximations*. New York: Springer-Verlag, 1997.

[9] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers, Part I: Unification and optimal designs," *IEEE Trans. Signal Processing*, vol. 47, pp. 1988–2006, July 1999.

[10] P. Stoica and B. C. Ng, "On the Cramér-Rao bound under parametric constraints," *IEEE Signal Processing Lett.*, vol. 5, pp. 177–179, July 1998.