RESOLUTION LIMITS OF BLIND MULTI-USER MULTI-CHANNEL IDENTIFICATION SCHEMES — THE BANDLIMITED CASE

Alle-Jan van der Veen

Delft University of Technology, Dept. Electrical Engineering/DIMES, 2628 CD Delft, The Netherlands

Blind space-time equalization and separation of multi-user digital communication signals presumes that the number of antennas M and the oversampling rate P is sufficiently large to be able to detect the number of sources and all channel lengths, and that the channel matrix is sufficiently well conditioned to allow inversion. A singular value analysis of the channel matrix for bandwidth limited signals provides necessary conditions for sufficient resolution, and guidelines for the selection of suitable M, P and equalizer lengths in relation to the bandwidth.

1. INTRODUCTION

A timely application area in signal processing is wireless (mobile) communications. We consider a scenario where several cochannel users are trying to talk to a central base-station over channels with large delay spread. In this case, there is both intersymbol interference and cochannel interference, requiring the use of multiple receiver antennas and space-time equalizers. Mathematically, the scenario is described as FIR-MIMO: finite impulse responses, multiple input signals (sources), multiple outputs (receivers). Several blind identification algorithms have been derived to solve individual aspects of the FIR-MIMO problem, in particular the more recent *subspace-based* approaches, that exploit the cyclostationarity property of digital signals by means of fractional sampling, and separate the signals based on their finite alphabet property [1–5].

One aspect of the problem that is independent of the actual algorithm is that of *resolution*: how many antennas and how much oversampling is needed to be able to detect the number of signals and estimate all channel lengths. There is not a single answer to this question. Generically, we have derived that the condition for identifiability is that $MP > d$, where M is the number of antennas, P the oversampling rate, and d the number of sources [3]. However, for bandlimited signals (as is likely the case in wireless RF communications), the role played by oversampling is limited: P and M are not equivalent any more. In this paper, we derive an expression that predicts the minimal number of antennas needed to separate and equalize a certain number of sources, as a function of the excess bandwidth, and assuming a large angle spread.

2. DATA MODEL

We use the data model of [3] which is summarized below. An array of M sensors, with outputs $x_1(t),\cdots,x_M(t)$, receives d digital signals $s_1(t),\cdots,s_d(t)$ through independent channels $h_{ij}(t)$. Each impulse response $h_{ij}(t)$ is a convolution of the shaping filter of the i-th signal and the actual channel from the i-th input to $x_j(t)$, including propagation delays and fractional delays necessary because signals need not be synchrounous. The data model is written compactly as the convolution $\mathbf{x}(t) = H(t) * \mathbf{s}(t)$, where

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_M(t) \end{bmatrix}, \; H(t) = \begin{bmatrix} h_{11}(t) & \cdots & h_{1d}(t) \\ \vdots & & \vdots \\ h_{M1}(t) & \cdots & h_{Md}(t) \end{bmatrix}, \; \mathbf{s}(t) = \begin{bmatrix} s_1(t) \\ \vdots \\ s_d(t) \end{bmatrix}$$

For a normalized symbol period ($T = 1$), assume that all $h_{ij}(t)$ are FIR filters of length at most $L \in \mathbb{N}$. Each $x_i(t)$ is sampled at a rate $P \in \mathbb{N}$, where P is the oversampling factor. Starting at time $t = 0$, and collecting samples during N symbol periods, we can construct a data matrix X as

$$X = [\mathbf{x}_0 \;\; \cdots \;\; \mathbf{x}_{N-1}] := \begin{bmatrix} \mathbf{x}(0) & \mathbf{x}(1) & \cdots & \mathbf{x}(N-1) \\ \mathbf{x}(\frac{1}{P}) & \mathbf{x}(1+\frac{1}{P}) & & \cdot \\ \vdots & & & \vdots \\ \mathbf{x}(\frac{P-1}{P}) & \cdot & \cdots & \mathbf{x}(N-1+\frac{P-1}{P}) \end{bmatrix}.$$

X has a factorization

$$X = HS_T = \begin{bmatrix} H(0) & H(1)\cdots H(L-1) \\ H(\frac{1}{P}) & \cdot \quad \cdot \\ \vdots & \vdots \\ H(\frac{P-1}{P}) & \cdot \quad \cdots H(L-1+\frac{P-1}{P}) \end{bmatrix} \begin{bmatrix} \mathbf{s}_0 & \ddots & \mathbf{s}_{N-2} & \mathbf{s}_{N-1} \\ \ddots & \ddots & \ddots & \mathbf{s}_{N-2} \\ \mathbf{s}_{-L+2} & \mathbf{s}_{-L+3} & \cdot^{\cdot^{\cdot}} & \ddots \\ \mathbf{s}_{-L+1} & \mathbf{s}_{-L+2} & \ddots & \mathbf{s}_{N-L} \end{bmatrix}$$

$$H:\; MP\times dL, \quad S_T:\; dL\times N, \text{ block-Toeplitz}. \tag{1}$$

The blind identification problem is to estimate H and S_T from X. Note that for such a factorization to be unique, it is necessary that H and S_T have full column rank and row rank, respectively, which implies a.o. $MP \geq dL$. If this condition does not hold, we can extend X to a block-Hankel matrix, by left-shifting and stacking m times,

$$\mathcal{X} = \begin{bmatrix} \mathbf{x}_0 & \mathbf{x}_1 & \cdot^{\cdot^{\cdot}} & \mathbf{x}_{N-m} \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} \\ \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \mathbf{x}_{N-2} \\ \mathbf{x}_{m-1} & \cdot^{\cdot^{\cdot}} & \mathbf{x}_{N-2} & \mathbf{x}_{N-1} \end{bmatrix} :\; mMP\times(N-m+1).$$

The augmented data matrix $\mathcal{X}$ has a factorization

$$\mathcal{X} = \mathcal{H}\mathcal{S} = \begin{bmatrix} \mathbf{0} & \boxed{H} \\ & \cdot^{\cdot^{\cdot}} \\ \boxed{H} & \\ \boxed{H} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{s}_{m-1} & \ddots & \mathbf{s}_{N-2} & \mathbf{s}_{N-1} \\ \ddots & & \ddots & \mathbf{s}_{N-2} \\ \mathbf{s}_{-L+2} & \mathbf{s}_{-L+3} & \ddots & \ddots \\ \mathbf{s}_{-L+1} & \mathbf{s}_{-L+2} & \ddots & \mathbf{s}_{N-L-m+1} \end{bmatrix}$$

$$\mathcal{H}: mMP\times d(L+m-1), \text{ block-Hankel},$$
$$\mathcal{S}: d(L+m-1)\times(N-m+1), \text{ block-Toeplitz}.$$

The stacking parameter m can be viewed as the length of an equalizer that tries to reconstruct $\mathcal{S}$ by forming linear combinations of the m block rows of $\mathcal{X}$. Necessary conditions for $\mathcal{X}$ to have a unique factorization $\mathcal{X} = \mathcal{H}\mathcal{S}$ are that $\mathcal{H}$ is a 'tall' matrix and $\mathcal{S}$ is a 'wide' matrix. The first condition leads to

$$MP > d, \qquad m \geq \frac{d(L-1)}{MP-d}. \tag{2}$$

$MP > d$ is a fundamental restriction. If $MP > d$, then we can always take m large enough to satisfy the second condition.

IEEE ICASSP'96, Atlanta (GA), May 1996.

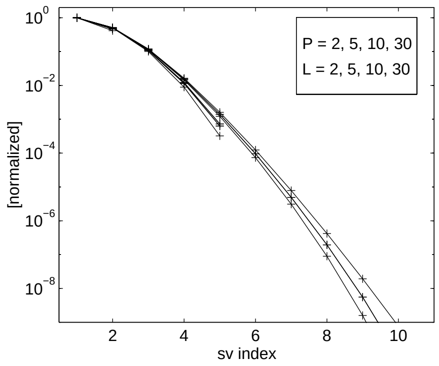

Figure 1. Singular values of Φ; $m = 1, \beta = 0$, varying P, L

Algorithms to find $\mathcal{H}$ and $\mathcal{S}$ from $\mathcal{X}$ under the condition that $\mathcal{H}$ has full column rank $d(L+m-1)$ were proposed in [1, 3], and extensions to unequal channel lengths in [2, 4, 5]. The effectiveness of these algorithms is limited by the conditioning of $\mathcal{H}$, which goes beyond the (practically useless) requirement of the "absence of common zeros" of the multidimensional channels.

3. BANDLIMITED SIGNALS

In view of Shannon's theorem, it would appear unlikely that it is possible to separate two bandwidth limited signals based on oversampling only: sampling beyond the Nyquist rate does not provide independent information. Typical communication signals use some excess bandwidth, *i.e.*, the Nyquist rate is larger but still close to the symbol rate. As a consequence, some information is gained by oversampling, but the role of P is limited, and $MP > d$ is not a sufficient condition to separate and equalize d signals.

If (2) holds and $\mathcal{H}$ and $\mathcal{S}$ have full rank, then $\text{rank}(\mathcal{X}) = d(L+m-1)$. As we show in this section, bandlimited signals generally lead to an ill-conditioned $\mathcal{H}$ and $\mathcal{X}$. Our objective is to derive minimal values for M and P in relation to the excess bandwidth β such that

1. a change in m by Δm increases the rank of $\mathcal{X}$ by $d\Delta m$,
2. a change in channel length L by ΔL increases the rank of $\mathcal{X}$ by $d\Delta L$.

Unless these two properties hold, we cannot expect any algorithm to provide good separation and equalization, since the number of signals and differences in channel lengths are not resolved.

3.1. One signal, one antenna

We start with the case where there is one signal and one antenna: $d = 1$, $M = 1$. A bandlimited signal is generated by a pulse shape function whose Fourier transform has only a limited number of non-zero coefficients, and since the channel is modeled as a linear system, the same holds for the convolution $h(t)$ of them. Let β represent the excess bandwidth, *i.e.*, the spectrum of the continuous-time signal is limited to $|f| \le (1+\beta)/2$. The block Hankel matrix $\mathcal{H}$ can be constructed from $[\mathbf{0}\ H]$ and cyclic shifts of it. Thus consider the augmented impulse response

$$h' = [\underbrace{0 \cdots 0}_{(m-1)P} \quad h_0 \quad h_{1/P} \quad \cdots \quad h_{L-1/P}],$$

which has length $L' := L+m-1$. The Fourier transform of h' has only $\alpha := L'(1+\beta)$ nonzero coefficients out of $L'P$, thus can be written as

$$h' = [f_1 \cdots f_\alpha] \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & \phi & \cdots & \phi^{L'P-1} \\ \vdots & \vdots & & \vdots \\ 1 & \phi^{\alpha-1} & \cdots & \phi^{(L'P-1)(\alpha-1)} \end{bmatrix}$$
$$\phi = \exp(\tfrac{j2\pi}{L'P}), \qquad \alpha = (L+m-1)(1+\beta).$$

A cyclic shift of h' leads to a cyclic shift of the columns of the DFT matrix, which can also be represented by premultiplying the DFT matrix with $\text{diag}[1, \phi^P, \cdots, \phi^{(\alpha-1)P}]$. After some manipulations, it follows that $\mathcal{H}$ can be factored as

$$\mathcal{H} = \Phi F V =$$
$$\left[\begin{array}{cccc} 1 & 1 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & \phi^{P-1} & \cdots & \phi^{(P-1)(\alpha-1)} \\ \hline 1 & \phi^{P} & \cdots & \phi^{P(\alpha-1)} \\ \vdots & \vdots & & \vdots \\ \hline \vdots & \vdots & & \vdots \\ 1 & \phi^{mP-1} & \cdots & \phi^{(mP-1)(\alpha-1)} \end{array}\right] \begin{bmatrix} f_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & f_\alpha \end{bmatrix} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & \phi^P & \cdots & \phi^{(L'-1)P} \\ \vdots & \vdots & & \vdots \\ 1 & \phi^{P(\alpha-1)} & \cdots & \phi^{(L'-1)P(\alpha-1)} \end{bmatrix}$$

$$\Phi : mP \times \alpha, \quad F : \alpha \times \alpha, \quad V : \alpha \times (L+m-1) \tag{3}$$

The rows of V are orthogonal, because they are full rows of a DFT(L')-matrix. F contains the non-zero channel Fourier coefficients, and we will assume in this analysis that it is not the limiting factor in the conditioning of $\mathcal{H}$, although, for $\beta > 0$, the coefficients are usually designed to taper off at the edges. Φ has dimensions $mP \times \alpha$ and is a principal submatrix of the DFT($L'P$) matrix. As a Vandermonde matrix, its conditioning can be quite bad, depending on m, P and β.

For example, suppose $\beta = 0$, $m = 1$, so that $\alpha = L$. The singular values of the corresponding Φ are plotted in figure 1 for a range of values for P and L. The objective is to see whether we can estimate L for cases where $P \ge L$: it was predicted by (2) that this is possible. The figure shows that, for $P \ge 2$, $L \ge 2$, the singular value plots are almost overlapping each other. The main effect of a larger P or L is that increasingly smaller singular values are added. For $L \ge 5$, say, we have to take such small singular values into account that the addition of only a tiny amount of noise (SNR around 60 dB) will already obscure these singular values and make the equalization fail. It is impossible to reliably estimate L.

For large m, Φ is a large submatrix of the full DFT matrix: its columns have length mP out of a total length of $(L+m-1)P$, and consequently, they are more independent of each other than was the case for $m = 1$. More precisely, one can prove that (for $\beta = 0$) Φ has a subset of m orthogonal columns, interleaved with $L-1$ other columns. Consequently, Φ has m large and approximately equal singular values. For general β, we obtain a similar result:

Proposition 1. *If Φ in (3) is a tall matrix, then it has $m(1+\beta)$ large and approximately equal singular values out of a total of $(L+m-1)(1+\beta)$.*

V is a tall matrix with orthonormal rows and reduces the dimension of ΦF from $\alpha \equiv (L+m-1)(1+\beta)$ columns to $L+m-1$. Grosso modo, the effect of multiplication by V can be modeled as a selection procedure which (statistically) retains the dominant $L+m-1$

singular values of ΦF. The model gets more reliable for larger reduction factors (here $1+\beta$). Since Φ and V are generated from the same DFT matrix, they are not independent, and this selection property is only true if F is sufficiently random. Note that F is generated by only $L(1+\beta)$ independent numbers (the nonzero Fourier coefficients of h), the other $(m-1)(1+\beta)$ nonzero entries are obtained by interpolation. Hence, there are limits to the effectiveness of a large m, and the above selection model fails once approximately $m > 2L$.

Proposition 1 allows to derive parameter values that are necessary for a good conditioning of $\mathcal{H}$ in the case of 1 antenna, 1 signal.

- Φ is a tall matrix if $mP \geq (L+m-1)(1+\beta)$, *i.e.*,

$$P > 1+\beta, \qquad m \geq \frac{L-1}{P-(1+\beta)}. \tag{4}$$

  To enable $m \leq 2(L-1)$, we should have $P \geq 1\frac{1}{2}+\beta$. There is no reason to take P much larger than that, as it will not improve the conditioning of $\mathcal{H}$.

- Only in case Φ has more large singular values than $\mathcal{H}$ has columns, $m(1+\beta) \geq L+m-1$, we can hope that all $L+m-1$ singular values of $\mathcal{H}$ are large. We refer to this as a "level 0" performance. It is equivalent to

$$m \geq \frac{L-1}{\beta} \qquad \text{[level 0]}. \tag{5}$$

This gives a minimal necessary condition on m. It may not be sufficient for detection of L. Note that $m(1+\beta) \geq L+m-1$ replaces the old condition $mP > L+m-1$: effectively, $P = 1+\beta$. To enable $m \leq 2(L-1)$, we should have $\beta \geq 0.5$. Simulations show that from that point on, changes in L affect the singular value plots of $\mathcal{H}$. For larger β the performance improves because the selection procedure by V is more reliable, but this is at the expense of an increased bandwidth. Simulations using raised-cosine pulse shape functions indicate that we need at least $\beta > 1$ for detection of L from a gap in singular values.

3.2. General singular value model for $\mathcal{H}$: $M \geq 1, d \geq 1$

With M antennas and d signals, we have a total of Md individual impulse responses. With some obvious rearrangements, the model for $\mathcal{H}$ is an extension of the model of section 3.1:

$$\mathcal{H} \sim \begin{bmatrix} \Phi F_{11} V \cdots \Phi F_{1d} V \\ \vdots \qquad \vdots \\ \Phi F_{M1} V \cdots \Phi F_{Md} V \end{bmatrix} = \underbrace{\begin{bmatrix} \Phi & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \Phi \end{bmatrix}}_{\Phi_M} \begin{bmatrix} F_{11} \cdots F_{1d} \\ \vdots \qquad \vdots \\ F_{M1} \cdots F_{Md} \end{bmatrix} \begin{bmatrix} V & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & V \end{bmatrix}$$

$$\Phi : mP \times \alpha, \quad F_{ij} : \alpha \times \alpha \text{ (diagonal)}, \quad V : \alpha \times (L+m-1)$$
$$\alpha \equiv (L+m-1)(1+\beta).$$

Φ_M is a tall matrix under the same conditions (4) as before. In that case, and assuming a large angle spread so that antennas give independent observations, Φ_M has approximately $Mm(1+\beta)$ large singular values. The Fourier coefficient matrix F selects the $d(L+m-1)(1+\beta)$ largest out of these, and the V-matrices further reduce this to the $d(L+m-1)$ largest singular values out of $d(L+m-1)(1+\beta)$.

The level-0 criterion is the requirement that Φ_M has more large singular values than are ultimately selected, *i.e.*,

$$Mm(1+\beta) \geq d(L+m-1) \tag{6}$$

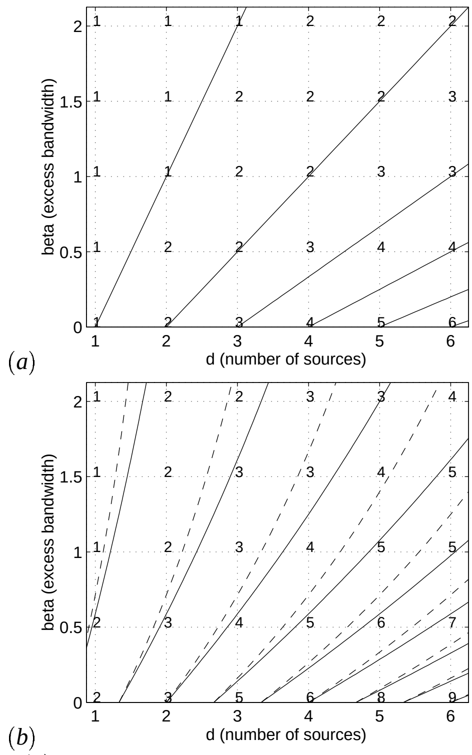

Figure 2. (a) Minimal values for the number of antennas M to detect the number of sources d (equation (7)), (b) Minimal values for M to detect changes in channel length L (equation (11), $p = 2$, $\varepsilon = 0.1$; dashed: $\varepsilon = 0.2$)

(Note that, again, P is effectively reduced to $P = 1+\beta$.) This implies, for $L > 1$,

$$M > \frac{d}{1+\beta} \qquad \text{[level 0]}. \tag{7}$$

This relation is plotted in figure 2(a). From simulations at each of the grid points in the figure, we have observed that it is a minimal condition on M such that increasing m by Δm increases the rank of $\mathcal{H}$ with $d\Delta m$, enabling detection of d, but perhaps not of L. With M satisfying (7), (6) gives conditions on m:

$$m \geq \frac{d(L-1)}{M(1+\beta)-d} \qquad \text{[level 0]}. \tag{8}$$

An improved "level-1 performance" is obtained when Φ_M has more large singular values than are selected by the F_{ij}, *i.e.*, $Mm(1+\beta) \geq d(L+m-1)(1+\beta)$, which implies

$$M > d, \qquad m \geq \frac{d(L-1)}{M-d} \qquad \text{[level 1]}. \tag{9}$$

Simulations using raised-cosine pulses indicate that level-1 performance usually gives a clear gap in singular values, enabling the estimation of L. An exception has to be made for small β ($\beta < 0.2$), because such signals have a very long impulse response of their own, requiring m to be very large. Especially when $M = d+1$, equation (9) might ask for $m \gg L$. As noted before, it does not make sense to take m much larger than $2L$, say, since data gets repeated and no new information is introduced. A performance somewhere between level 0 and 1 is such that a change of ΔL in L increases the rank of $\mathcal{H}$ by $d\Delta L$.

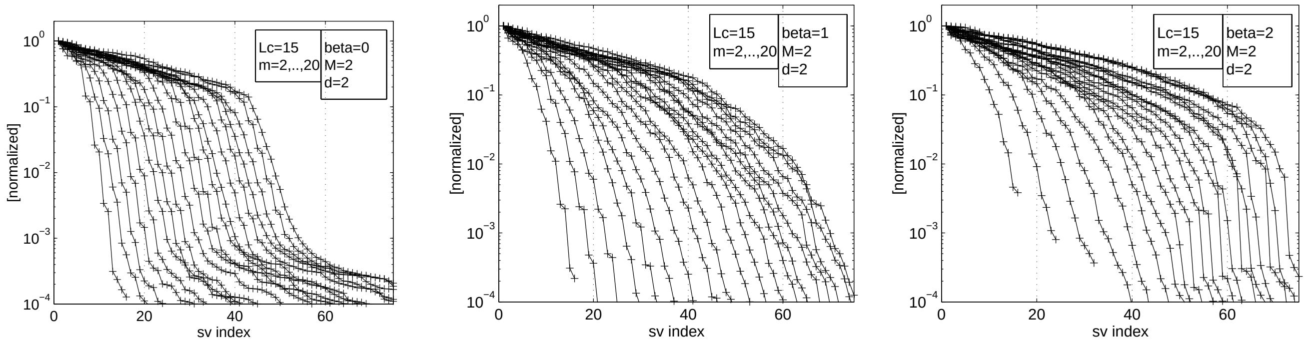

Figure 3. Singular value plots of $\mathcal{H}$ for varying m, 2 antennas, 2 signals. (a) $\beta = 0$, (b) $\beta = 1$, (c) $\beta = 2$.

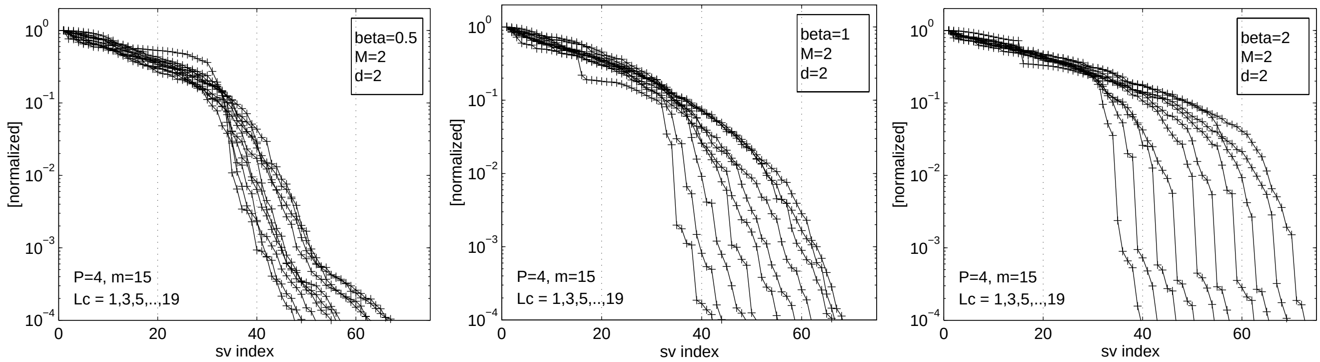

Figure 4. Singular value plots of $\mathcal{H}$ for varying L, 2 antennas, 2 signals. (a) $\beta = 0.5$, (b) $\beta = 1$, (c) $\beta = 2$.

4. SELECTION OF *P* AND *M*

We summarize the above conditions into criteria for the selection of the oversampling rate P and the number of antennas M. A performance level ε between 0 and 1 is obtained when

$$M(1+\beta)m \;\geq\; d(L+m-1)(1+\varepsilon\beta)\,, \qquad 0 \leq \varepsilon \leq 1\,. \qquad (10)$$

Further suppose that $m := p(L-1)$ where we will restrict p to $p \leq 2$. This reduces (4) and (10) to

$$P \;\geq\; 1 + \frac{1}{p} + \beta, \qquad M \;\geq\; d\frac{1+p}{p}\frac{1+\varepsilon\beta}{1+\beta}\,.$$

If we settle for $p = 2$, then we obtain

$$P \;\geq\; 1\tfrac{1}{2} + \beta, \qquad M \;\geq\; 1\tfrac{1}{2} d\frac{1+\varepsilon\beta}{1+\beta}\,. \qquad \text{[level } \varepsilon\text{]} \qquad (11)$$

Figure 2(b) shows this relation for $\varepsilon = 0.1$ and $\varepsilon = 0.2$ (dashed). Lines of constant M are hyperbolas in the graph. Note that the required number of antennas is linear in d. The obtained values of M should be regarded as minimal values, below which an increase of L by ΔL will not increase the rank of $\mathcal{H}$ by $d\Delta L$. To have a clear gap between the large and small singular values of $\mathcal{H}$ requires more: $\varepsilon \geq 0.5$ or so. For $\varepsilon = 0$, we can essentially only expect that d can be detected from changes in m, and that an increase of L has "some effect" in the singular value plots. For small L (say $L < 5$), the values of figure 2(a) are already sufficient.

As an example, figure 3 shows the singular values of $\mathcal{H}$ for simulated channels for $d = 2$ signals, $M = 2$ antennas, constant $L = 15$ and varying m. To detect d from variations in m, figure 2(a) predicts that we need $\beta \geq 0$, and indeed, even for $\beta = 0$ the rank increases as it should, although L cannot be estimated correctly. Figure 4 shows what happens when L is varied, for constant m. According to figure 2, we need at least $\beta > 0.75$ or so to observe an effect in changes of L. Indeed, for $\beta = 0$, the channel length cannot be determined at all: all lines overlap (plot omitted). For $\beta = 0.5$, some effect of changing L is seen, but not at all well-determined. For $\beta = 1$, it is possible to detect that $\Delta L = 2$, provided m is large enough in relation to L (as determined by (10)). For $\beta = 2$, the rank of $\mathcal{H}$ is clear and it becomes possible to estimate L itself as well. Backward calculation shows that this case has level $\varepsilon = 0.5$.

5. REFERENCES

[1] H. Liu and G. Xu, "A deterministic approach to blind symbol estimation," *IEEE Signal Processing Letters*, vol. 1, pp. 205–207, Dec. 1994.

[2] H. Liu and G. Xu, "Multiuser blind channel estimation and spatial channel pre-equalization," in *Proc. IEEE ICASSP*, (Detroit), pp. 1756–1759 vol.3, May 1995.

[3] A.J. van der Veen, S. Talwar, and A. Paulraj, "Blind estimation of multiple digital signals transmitted over FIR channels," *IEEE Signal Processing Letters*, vol. 2, pp. 99–102, May 1995.

[4] A.J. van der Veen, S. Talwar, and A. Paulraj, "Blind identification of FIR channels carrying multiple finite alphabet signals," in *Proc. IEEE ICASSP*, (Detroit), pp. 1213–1216, May 1995.

[5] A.J. van der Veen, S. Talwar, and A. Paulraj, "Blind estimation of multiple digital signals transmitted over multipath channels," in *Proc. IEEE MILCOM*, (San Diego), Nov. 1995.