INVERTING OVERDETERMINED TOEPLITZ SYSTEMS WITH APPLICATION TO BLIND BLOCK-ADAPTIVE EQUALIZATION

Anna Scaglione, Sergio Barbarossa

Infocom Dept., Univ. of Rome "La Sapienza",
via Eudossiana 18, 00184 Roma, ITALY
{sergio,annas}@infocom.ing.uniroma1.it

Georgios B. Giannakis

Dept. of Electrical Engr., Univ. of Virginia,
Charlottesville, VA 22903-2442, USA
e-mail: georgios@virginia.edu

ABSTRACT

Linear channel equalization in block transmission systems amounts to inverting Toeplitz systems of linear equations. Motivated by limitations of a recent blind block equalizer, we derive properties and investigate the class of tall Toeplitz matrix inverses which themselves exhibit (even approximate) Toeplitz structure. The class is characterized by the size of leading and trailing all-zero block submatrices, and interesting links as well as optimal choices of the size parameter are established with the number of maximum-phase zeros of the underlying channel transfer function. Exploiting the properties of such equalizers we derive a direct blind adaptive equalizer and illustrate superiority over competing approaches. It is also shown that the optimum delay for blind block equalization corresponds to the number of maximum-phase channel zeros.

1. INTRODUCTION

In a single input single output (SISO) system, perfect linear equalization of non-minimum phase FIR channels can be achieved with stable IIR (non-) causal filters, provided that the channel transfer function does not have zeros on the unit circle. Even if channel zeros are in the vicinity of the unit circle, the performance of linear equalizers degrades. On the other hand, channel equalization in block transmission systems, amounts to inverting Toeplitz systems of linear equations. Such a block transmission scheme was dealt with in [4], where it was shown that by adding redundancy to each block in the form of guard bits (as many as the channel order) guarantees existence of block equalizers *irrespective* of the channel zero locations. Guard intervals eliminate interblock interference (IBI), so the equalizer needs to cancel the intersymbol interference (ISI) only within a single block.

The work in this paper was supported by NSF-CCR grant no. 9805350.

Consider a block synchronous transmission system, and denote with $\boldsymbol{s}(n) := [s(nM) \ldots s(nM+M-1)]^T$ the $M \times 1$ channel input block; with $\boldsymbol{H}$ the $(M+L) \times M$ the Toeplitz channel matrix, having $[h(0) \ldots h(L) 0 \ldots 0]$ as its first column, and $[h(0) 0 \ldots 0]$ as its first row; and with $\boldsymbol{x}(n) := [x(nP) \ldots x(nP+P-1)]^T$ the $P \times 1$ channel output block, where L is the channel order and $P := M+L$. If we choose $M > L$, then we can express the corresponding noise free output data block as:

$$\boldsymbol{x}(n) = \boldsymbol{H}\boldsymbol{s}(n) \quad . \tag{1}$$

From (1), the block equalizer is well defined because the system of linear equations is *always* invertible and the solution that minimizes the error norm is given by $\boldsymbol{G} = \boldsymbol{H}^{\dagger}$, where $\dagger$ denotes matrix pseudoinverse. Intuitively, even deep fades can be equalized because the presence of guard bits allows one to equalize the channel by solving an overdetermined system of linear equations. In fact, each block of M input symbols is mapped to a block of $P = M+L$ data. However, equalization requires knowledge of the channel which may not be available (or bandwidth-consuming to acquire) in wireless communications systems. For such systems, a deterministic (semi-)blind scheme was developed recently in [4], that obviates channel estimation and directly obtains the block equalizer by exploiting guard intervals in the form of distributed training. However, it was observed in [4] that the resulting direct equalizer tends to emphasize noise effects when the channel is non-minimum phase. Motivated by the limitations of the blind block equalizer in [4] this paper explores reasons behind and remedies for this undesired behavior. A blind adaptive equalizer is also developed which, contrary to the one in [4], it has performance close to the minimum mean square error (MMSE) ZF solution with known channel.

2. OVERDETERMINED TOEPLITZ

Because the blind equalizer $\boldsymbol{G}$ in [4] has a lower triangular Toeplitz form, we first derive properties of $\boldsymbol{H}$

inverses that exhibit Toeplitz or approximate Toeplitz structure. Specifically, we establish that:

- if the channel is non-minimum phase, among all the possible square matrices inverting equation (1), the direct blind equalization method in [4] leads to a solution G_0 that is a truncated unstable zero-forcing (ZF) equalizer;
- the equalizer G_0 is one of the $L+1$ possible equalizers $\{G_d\}_{d=0}^{L}$, that are Toeplitz, or have approximate Toeplitz structures, and can be obtained by inverting the matrix composed of the d^{th} row of H up to the $(M+d)^{th}$. If we let $H(z) := \mathcal{Z}\{h(n)\}$ denote the channel transfer function, the "best" (with respect to numerical robustness) set of equations corresponding to the minimum matrix norm[1] $||G_d||$, is obtained by selecting $d = d_{opt}$, where d_{opt} is the number of zeros ρ_l of $H(z)$ outside the unit circle;
- for $M \gg L$, the rows of G_d are approximately truncated versions of stable ZF equalizers if and only if $d = d_{opt}$.

Solving square systems of equations corresponding to different delays d, provides an insight to selecting delays in linear equalizers [3]. For a given FIR channel with no zeros on the unit circle, it is well known that there exists a delay-d SISO equalizer $G_d(z)$ such that:

$$H(z)G_d(z) = z^{-d} \quad \text{with} \quad \sum_{n=-\infty}^{\infty} |g_d(n)|^2 < \infty \,. \quad (2)$$

Because d is not unique, the delay should be selected optimally to minimize, for example, the noise power at the equalizer's output. Let us introduce the following notation:

$$I_d := \begin{pmatrix} 0_{d\times M} \\ I_{M\times M} \\ 0_{L-d\times M} \end{pmatrix} \; J := \begin{pmatrix} 0 & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 0 & \dots & \dots & 0 \end{pmatrix}, \quad (3)$$

where I_d has dimensionality $(M+L)\times M$ and J is a square shift matrix. Since J in not invertible, we will use the notation J^{-1} to denote:

$$J^{-1} := J^{\dagger} \equiv J^T \quad (4)$$

where ^{T} denotes transpose. The counterpart of (2) for block transmissions, and the problem we address boils down to:

$$\min_d ||G_d|| \quad \text{subject to} \quad G_d I_d^T H = I_{M\times M} \,. \quad (5)$$

[1] Any matrix norm $||.||$ satisfying the submultiplicative property $||AB|| \leq ||A||||B||$ can be considered.

Since $I_d^T H$ is a square matrix, the ZF constraint imposes that $G_d = (I_d^T H)^{-1}$, and so

$$\min_d ||G_d|| = \min_d ||(I_d^T H)^{-1}||. \quad (6)$$

The optimal d can be obtained by first decomposing matrix H as:

$$H = h(0)I_0 + h(1)I_1 + \ldots + h(L)I_L. \quad (7)$$

It is straightforward to verify that $I_j^T I_i = J^{j-i}$; hence, we can write the following factorization of $I_d^T H$:

$$\begin{aligned} I_d^T H &= h(0)J^d + h(1)J^{d-1} + \ldots + h(L)J^{d-L} \\ &= h(0)J^d \prod_{l=1}^{L}(I - \rho_l J^{-1}). \end{aligned} \quad (8)$$

Similar to $z^d H(z) = h(0)z^d \prod_{l=1}^{L}(1-\rho_l z^{-1})$, considering that approximately, for $M \gg 1$, $I \simeq JJ^{-1}$ (because $\text{diag}(JJ^{-1}) = [\underbrace{1\ldots 1}_{M-1}, 0]$), we can write:

$$I_d^T H \simeq h(0)\prod_{l=1}^{d}(-\rho_l)(I - \rho_l^{-1}J) \prod_{l=d+1}^{L}(I - \rho_l J^{-1})\,. \quad (9)$$

The approximation in (9) is valid in the sense that the ratio of the error norm over the matrix norm $||I_d^T H||$ tends to zero as M increases. From (9), we infer that $G_d = (I_d^T H)^{-1}$ can be factorized as:

$$\begin{aligned} G_d &\simeq \frac{(-1)^d}{h(0)}\prod_{l=1}^{d}\frac{1}{\rho_l}(I - \frac{1}{\rho_l}J)^{-1}\prod_{l=d+1}^{L}(I - \rho_l J^{-1})^{-1} \quad (10) \\ &= \frac{(-1)^d}{h(0)}\left[\prod_{l=1}^{d}\sum_{m=0}^{M-1}\frac{1}{\rho_l^{m+1}}J^m\right]\left[\prod_{l=d+1}^{L}\sum_{m=0}^{M-1}\rho_l^m J^{-m}\right], \end{aligned}$$

which implies that $(I_d^T H)^{-1}$ tends to a Toeplitz matrix. On the other hand, any Toeplitz matrix is liable of this decomposition, which is not possible if we interchange the equations selected in any of these L sets corresponding to the matrix of coefficients $I_d^T H$. Therefore, we are dealing with the class of solutions G_d of (1) that are (perhaps approximate) Toeplitz matrices.

Based on (10) we establish the following:

Lemma 1: *The optimal solution of* (5) *is given by* $\mathbf{G}_d = (I_d^T H)^{-1}$ *where* $d = d_{opt}$, *with* $d_{opt} :=$ *number of zeros* ρ_l *of* $H(z)$ *with* $|\rho_l| > 1$. *In particular, we prove that:*

$$\lim_{M\to\infty}\frac{1}{M}||(I_d^T H)^{-1}|| = \begin{cases} \infty & d \neq d_{opt} \\ 0 & d = d_{opt} \end{cases} \quad \square \quad (11)$$

Interestingly, this Lemma provides as a by-product a blind deterministic method for identifying the number of channel zeros outside the unit circle – a task impossible in the standard SISO setup with output second-order statistics. Specifically, if we collect a data matrix $X_N = HS_N$ with $N > M+L$, then:

$$d_{opt} = \text{argmin}_d ||(I_d^T X_N)^{-1}||. \quad (12)$$

3. BLIND ADAPTIVE EQUALIZER

The system model adopted in [4] is more general than (1), because it includes a filterbank precoder. The precoder performs a linear mapping $\boldsymbol{F}$ of $\boldsymbol{s}(n)$ and yields the following I/O relationship:

$$\boldsymbol{x}(n) = \boldsymbol{H}\boldsymbol{F}\boldsymbol{s}(n). \tag{13}$$

We assume for simplicity that $\boldsymbol{F} = \boldsymbol{I}$ that allow us to use (1) directly . However, the results in this section apply to arbitrary $\boldsymbol{F}$ by multiplying the equalizer herein by $\boldsymbol{F}^{-1}$ from the left.

The ideal direct blind equalizer $\hat{\boldsymbol{G}}$ derived in [4] corresponds to the member $\boldsymbol{G}_0$ of the class in (5). Notice that for $d = 0$ and $d = L$ both (9) and (10) are *exact.* Hence, $\boldsymbol{G}_0$ ($\boldsymbol{G}_L$) is Toeplitz and lower (upper) triangular. The idea underlying the method in [4] is to exploit the specific structure of $\boldsymbol{G}_0$ to estimate its last row $\boldsymbol{\gamma}_M^H$. Because $\boldsymbol{G}_0$ is Toeplitz and lower triangular, the last row specifies every other row of $\boldsymbol{G}_0$. Indeed, denoting by $\boldsymbol{\gamma}_i^H$ the ith row of $\boldsymbol{G}_0$, we have that

$$\boldsymbol{\gamma}_i^H = \boldsymbol{\gamma}_{i+1}^H \boldsymbol{J}^{-1} \quad i = 1, \ldots, M. \tag{14}$$

More specifically, let us define the $P \times N$ data matrix $\boldsymbol{X}_N = (\boldsymbol{x}(n) \ \ldots \ \boldsymbol{x}(n(N-1)))$, with $N > P$. It was shown in [4] that the $P \times 1$ vector $\boldsymbol{\gamma}$, satisfying $\boldsymbol{\gamma}_M = \boldsymbol{I}_L^T \boldsymbol{\gamma}$, can be identified as the unique solution (within a scale) of:

$$\boldsymbol{\gamma}^H [\boldsymbol{X}_N \, \boldsymbol{J}^{-1}\boldsymbol{X}_N \ \ldots \boldsymbol{J}^{-L+1}\boldsymbol{X}_N] = \boldsymbol{0}\,. \tag{15}$$

Hence, $\boldsymbol{G}_0$ can be also identified up to a scalar factor, using (14). First, we note that with a slight modification the method in [4] yields also $\boldsymbol{G}_L$. In particular it can be proved, following the same lines as in [4], that $\boldsymbol{G}_L$ can be identified up to a scalar factor by solving:

$$\boldsymbol{\beta}^H [\boldsymbol{X}_N \, \boldsymbol{J}\boldsymbol{X}_N \ \ldots \boldsymbol{J}^{L-1}\boldsymbol{X}_N] = \boldsymbol{0}\,, \tag{16}$$

where $\boldsymbol{\beta}$ is a $P \times 1$ vector, such that the first row $\boldsymbol{\beta}_1^H$ of $\boldsymbol{G}_L$ is $\boldsymbol{\beta}_1 = \boldsymbol{I}_0^T \boldsymbol{\beta}$ and every other row can be obtained as $\boldsymbol{\beta}_{i+1}^H = \boldsymbol{\beta}_i^H \boldsymbol{J}$. Lemma 1 implies that $\boldsymbol{G}_L$ enhances the noise whenever there are channel zeros *inside* the unit circle. Hence, even if both $\boldsymbol{G}_0$ and $\boldsymbol{G}_L$ estimates are consistent, the average (over several channels) performance in terms of bit error rate of both equalizers will be far from the performance of the MMSE ZF equalizer given by $\boldsymbol{G}_{mmse-zf} = \boldsymbol{H}^\dagger$.

To overcome this undesired effect and offer reliable equalizers irrespective of the channel zero locations, we propose a novel adaptive blind algorithm that estimates $\boldsymbol{G}_{d_{opt}}$ from the noisy data $\boldsymbol{Y}_N := \boldsymbol{X}_N + \boldsymbol{V}_N$, where $\boldsymbol{V}_N$ is AGN. The price paid is an increase of computational complexity with respect to the algorithm that estimates $\boldsymbol{G}_0$ or $\boldsymbol{G}_L$ alone. The method is based on the following equation, which allows us to construct *any* $\boldsymbol{G}_d$ from the noise-free data block $\boldsymbol{X}_N$ and the estimates of $\boldsymbol{G}_0$ or $\boldsymbol{G}_L$:

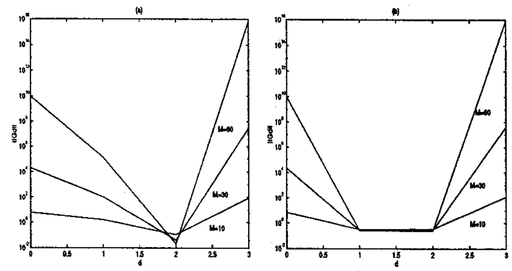

Figure 1: $(1/M)||\boldsymbol{G}_d||_2$ versus d.

$$\boldsymbol{G}_d \boldsymbol{I}_d^T \boldsymbol{X}_N = \boldsymbol{G}_0 \boldsymbol{I}_0^T \boldsymbol{X}_N = \boldsymbol{G}_L \boldsymbol{I}_L^T \boldsymbol{X}_N, \tag{17}$$

hence:

$$\begin{aligned} \boldsymbol{G}_{d_{opt}} &= \boldsymbol{G}_0 \boldsymbol{I}_0^T \boldsymbol{X}_N (\boldsymbol{I}_{d_{opt}}^T \boldsymbol{X}_N)^\dagger \\ &= \boldsymbol{G}_L \boldsymbol{I}_L^T \boldsymbol{X}_N (\boldsymbol{I}_{d_{opt}}^T \boldsymbol{X}_N)^\dagger \end{aligned} \tag{18}$$

In the context of direct blind fractionally spaced equalizers, related optimum delay selection ideas were discussed in [1].

Denoting the noisy data by $\boldsymbol{y}(n) = \boldsymbol{x}(n) + \boldsymbol{v}(n)$, the algorithm consists of the following steps:

1. Collect $N > P$ data blocks in a matrix $\boldsymbol{Y}_N^{(0)} := (\boldsymbol{y}(0) \ \ldots \ \boldsymbol{y}(N-1))$, estimate d_{opt} as in (12) and set $i = 0$, $\boldsymbol{Y}_N^{(i)} := (\boldsymbol{y}(i) \ \ldots \ \boldsymbol{y}(i+N-1)) = \boldsymbol{Y}_N^{(0)}$ and $\hat{\boldsymbol{G}}_{d_{opt}}^{(0)} = \boldsymbol{0}$.
2. Estimate $\hat{\boldsymbol{G}}_0^{(i)}$ and $\hat{\boldsymbol{G}}_L^{(i)}$ using (15) and (16).
3. Estimate $\hat{\boldsymbol{G}}_{d_{opt}}^{(i)}$ as follows [cf. (18)]:

$$\begin{aligned} \hat{\boldsymbol{G}}_{d_{opt}}^{(i)} &= \lambda(i)\hat{\boldsymbol{G}}_{d_{opt}}^{(i-1)} + \left[\left(1 - \frac{d_{opt}}{L}\right)\hat{\boldsymbol{G}}_0^{(i)} \boldsymbol{I}_0^T \boldsymbol{X}_N \right. \\ &+ \left. \frac{d_{opt}}{L}\hat{\boldsymbol{G}}_L^{(i)} \boldsymbol{I}_L^T \boldsymbol{X}_N \right] (\boldsymbol{I}_{d_{opt}}^T \boldsymbol{X}_N)^\dagger \end{aligned} \tag{19}$$

   where $\lambda(i)$ is a suitable forgetting factor depending on the iteration index.
4. Update matrix $\boldsymbol{Y}_N^{(i+1)}$ with the new observation $\boldsymbol{y}(i+N)$.
5. Set $i = i + 1$ and go back to step 2.

4. NUMERICAL RESULTS

We present next simulations to validate our theory.
Example 1: Fig. 1 shows the values of $(1/M)||\boldsymbol{G}_d||_2 = (1/M)\ ||(\boldsymbol{I}_d^T\boldsymbol{H})^{-1}||_2$ versus d for block lengths $M = 10, 50, 100$ and a channel of order $L = 4$ with roots: a) (left) $(1.3, j0.5, 0.5-j1.5, -0.5)$; b) (right) $(1, j0.5, 0.5-j1.5, -0.5)$. It is interesting that in case (a) the value

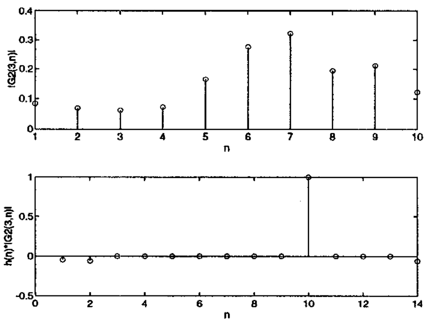

Figure 2: Third row of G_2 for the channel of Fig. 1(a) ($d_{opt} = 2$, $M = 10$), and its convolution with the channel.

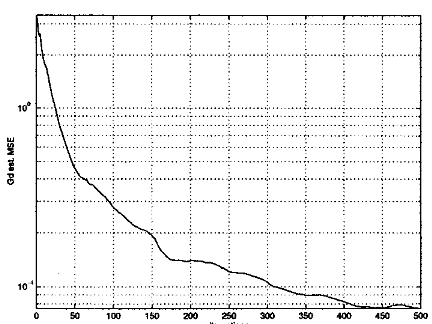

Figure 3: $\hat{G}^{(i)}_{d_{opt}}$ MSE vs. number of iterations.

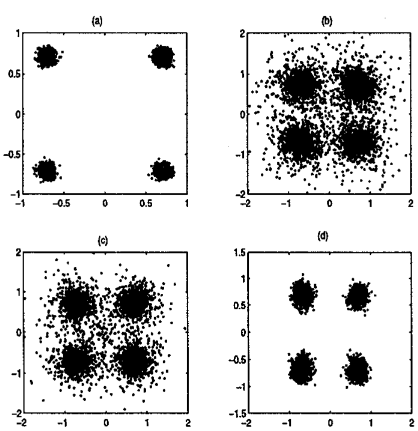

Figure 4: Scatter diagram: a) $G_{mmse-zf}$ with known channel, b) $\hat{G}_0^{(500)}$; c) $\hat{G}_L^{(500)}$; d) $\hat{G}^{(500)}_{d_{opt}}$.

of $(1/M)||G_d||_2$ vs. d has an evident minimum at $d = d_{opt}$, in contrast to case (b) where a unit circle zero results in one or two minima of the curve $(1/M)||G_d||_2$ depending on how one considers the boundary.

Example 2: Fig. 2 shows the convolution of the 3rd row of G_2 with the same channel of Example 1 (left graph of Fig. 1 a), which tends to a delta function. This validates our interpretation of the block inverse G_d as an approximation of the Toeplitz matrix corresponding to the truncated version of the IIR channel inverse impulse response $z^{-d}/H(z)$.

Example 3: Fig. 3 depicts the mean square error (MSE) $||G_{d_{opt}} - \hat{G}_{d_{opt}}||$ of the equalizer estimate vs. the number of iterations. The step size used at the ith iteration is $\lambda(i) = 0.6/i$, $M = 16$, and the energy per symbol E_s over the noise spectral density N_0 is $E_s/N_0 = 26$ dB. The channel has $L = 4$ zeros at $[j1.2, -j1.2, -0.6(1+j), 0.6(1+j)]$ which, according to Lemma 1, implies $d_{opt} \equiv 2$, as was correctly found form (12) with $Y_N = Y_N^{(0)}$. Note that E_s/N_0 is quite low relative to blind CMA equalizer requires. This is due to the fact that both the estimates $\hat{G}_0$ and $\hat{G}_L$ are consistent (see [4]), i.e. for $N \to \infty$ the estimates converge to the true equalizers, which guarantees also consistency of $\hat{G}_{d_{opt}}$.

The four scatter diagrams in Fig. 4 are obtained from the same simulation as in Fig. 3, using: i) the MMSE-ZF equalizer derived from the true channel (top left); ii) the equalizer $\hat{G}_0$ (top right); iii) the equalizer $\hat{G}_L$ (bottom left); iv) the equalizer $\hat{G}_{\hat{d}_{opt}}$ (bottom right), where equalizer estimates correspond to $i = 500$ and the last iteration of our adaptive algorithm. In agreement with theory, the equalizer approaches the MMSE-ZF equalizer performance.

5. REFERENCES

[1] G. B. Giannakis and C. Tepedelenlioğlu, "Direct blind equalizers of multiple FIR channels: A deterministic approach," *IEEE Transactions on Signal Processing*, vol. 47, January 1999 (to appear).

[2] G. H. Golub, C. F. Van Loan "Matrix computation", pp.79-81, *The Johns Hopkins University Press*, 1989.

[3] A. Erdogan, B. Hassibi, T. Kailath "H-infinity equalization of communication channels." *Proc. ICASSP '98*, vol. VI, pp. 3433-3437, Seattle, May 1998.

[4] A. Scaglione, G. B. Giannakis, S. Barbarossa, "Self-recovering multirate equalizers using redundant filter-bank precoders," *Proc. of Intl. Conf. on ASSP*, vol. VI, pp. 3501-3504, Seattle, WA, May 12-15, 1998.