ARCC Workshop on *Time-reversal communications in richly scattering environments*

Large-Delay-Spread Channels: Mathematical models and fast algorithms

Thomas Strohmer

Department of Mathematics

University of California, Davis, USA

strohmer@math.ucdavis.edu

http://math.ucdavis.edu/~strohmer

Talk includes joint work with Arogyaswami Paulraj, Majid Emami (Stanford), Jan Hansen (Stanford/ETH), Raymond Guan (UC Davis)

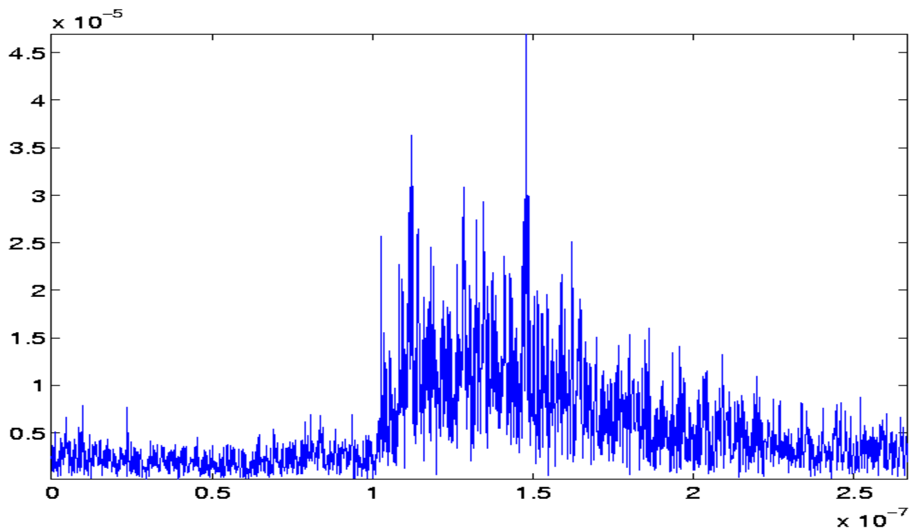
Research supported by NSF DMS

Agenda

- Challenges of HDB channels
- A mathematical framework for equalization
- Pre/post/joint equalization
- Krylov subspace methods + HDB channels

What makes HDB channels difficult?

- large AWGN (due to large bandwidth, interfering devices,...)
- low signal power (due to FCC regulations, equipment limitations,...)
- large delay spread
- deep fading channels
- channel estimation
- Doppler spread
- interfering users
- low computing power/storage
- request for high data rates
- ...


x 10^-5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0.5
1
1.5
2
2.5
x 10^-7

What is different when delay spread and bandwidth are large?

- Current equalization methods (ML, even MMSE) can become prohibitively expensive for large delay spread channels (despite great progress such as Babak Hassibi's work on spherical decoder).

- Channel is more likely to have deep fades

- Finite-dimensional approximations of properties valid only in infinite dimensions becomes less justified (e.g., inverse of biinifinite Toeplitz matrix is Toeplitz, but inverse of finite Toeplitz matrix is not Toeplitz)

Goal: Simple equalizer

Several ways to define "simple equalizer":

(i) need to estimate only few taps of (pre-coded) channel impulse response h

(ii) equalizer itself has only few taps, but h may still have many taps, e.g. FIR-equalizer with few taps

(iii) equalizer is computationally cheap

(iv) combination of any of the above

Brute-force approaches to "reduce" delay spread: truncate channel impulse response h or use only N largest taps of h.

Matrices and equalization

$x = \{x_{k,n}\}_{k\in\mathbb{Z}}$ data symbols to be transmitted,

$$\varphi_{k,n}(t) = \varphi_n(t - kT), \quad n = 0, \ldots, N-1$$

are the transmission pulses, $\mathbf{H}$ is channel.

$N = 1$: single-carrier system
$N > 1$, $\varphi_n(t) = \varphi(t)e^{2\pi itnF}$: multicarrier system

Define operator C by

$$Cx := \sum_{k,n} x_{k,n}\varphi_n(t - kT)$$

Transmit signal is $s = Cx$, received signal is $r = \mathbf{H}s + \eta$. Assume $\eta = 0$ for now. Note that

$$C^*s = \{\langle s, \varphi_{k,n}\rangle\}_{k,n}.$$

After applying matched filter C^* at receiver coefficients $y = \{y_{k,n}\}$ are given by

$$y_{k,n} = (C^* r)_{k,n} = \langle \mathbf{H} s, \varphi_{k,n} \rangle = \\ = \sum_k x_{k',n'} \langle \mathbf{H} \varphi_{k',n'}, \varphi_{k,n} \rangle$$

Recall:

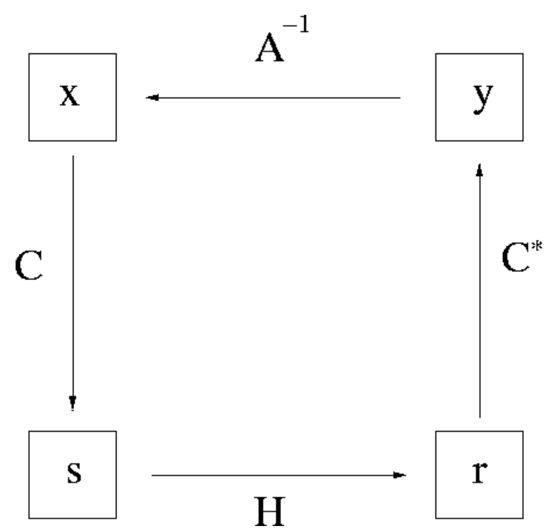
$$y_{k,n} = \sum_{k',n'} x_{k',n'} \langle \mathbf{H}\varphi_{k',n'}, \varphi_{k,n} \rangle$$

Define matrix A

$$A_{k,n,k',n'} = \langle \mathbf{H}\varphi_{k',n'}, \varphi_{k,n} \rangle.$$

Then

$$Ax = C^*\mathbf{H}Cx = y$$



Let Ω be transmission bandwidth. For A to be invertible we need that $\{\varphi_{k,n}\}$ are linearly independent. For optimal spectral efficiency we need that $\{\varphi_{k,n}\}$ is an orthonormal basis for space of Ω-bandlimited signals.

Efficient equalization: want A to be diagonal, but need also take into account costs for applying operators C, C^*.

OFDM

Channel impulse response $h = [h_0, \ldots, h_{L-1}]$. Consider OFDM system with N tones and guard interval or cyclic prefix of length $\geq L$.

ISI is avoided by using cyclic prefix, matrix A is diagonal.

Equalization is done via FFTs, computational costs for CP-OFDM: $\mathcal{O}(N \log N)$.

But for large delay spread cyclic prefix becomes very long (as for any block transmission scheme) and coherence bandwidth very short: loss of spectral efficiency, large number of tones needed

However if N is very large, block-fading model of channel may no longer apply.

Other problems of OFDM: large peak-to-average ratio of transmission signals, sensible to carrier offsets

Example-Theorem [T.S. 2004]: Assume $\mathbf{H}$ is a channel with delay spread and Doppler spread (for mathematicians: $\mathbf{H}$ is a pseudodifferential operator whose Weyl symbol belongs to weighted Sjöstrand class). Assume bandwidth is not "too large".

Then A is almost diagonal if and only if transmission pulses of multicarrier system are of the form

$$\varphi_{k,n}(t) = \varphi(t - kT)e^{2\pi inFt}$$

where φ is a transmission pulse whose localization in time and frequency depend on amount of delay spread and Doppler spread.

Meaning of "almost diagonal" and "localization of φ" can be made mathematically rigorous.

φ can be constructed explicitly

From now on we assume only delay spread and AWGN but no Doppler spread.

We also assume that design of transmission pulses has already been carried out.
E.g. $\varphi =$ raised cosine.

Thus for simplicity of notation we assume that impulse response h comprises tranmission pulses and actual channel.

Matrices with off-diagonal decay and their inverses

Mathematical backbone for equalizer design with few taps

Simple case: band matrices

Assume A is an m-band matrix, i.e.,

$$A_{k,l} = 0, \qquad \text{if } |k-l| \geq m.$$

Let $\kappa = \text{cond}(A) := \|A\|_2 \|A^{-1}\|_2$. If A is invertible then

$$|A^{-1}_{k,l}| \leq c\lambda^{|k-l|/m}, \qquad \lambda < 1$$

where λ and c depend only on κ.

Increasing κ means $\lambda \to 1$ and increasing C.

If impulse response h has L taps then channel matrix H is (non-symmetric) L-band matrix. If L is large, decay estimate for inverse may be very pessimistic, since we did not take into account (slow) exponential decay of taps.

Sketch of general theory:[Baskakov '90, '97] (based on noncommutative Banach algebras)

Let $A = [A_{k,l}]_{k,l \in \mathcal{I}}$ be a matrix from $\ell^2(\mathcal{I})$ to $\ell^2(\mathcal{I})$ where $\mathcal{I} = \mathbb{Z}$ or some other index set. Assume that

$$|A_{k,l}| \le d_\alpha(|k-l|)$$

where $d_\alpha(x)$ is a "decaying" function, such as

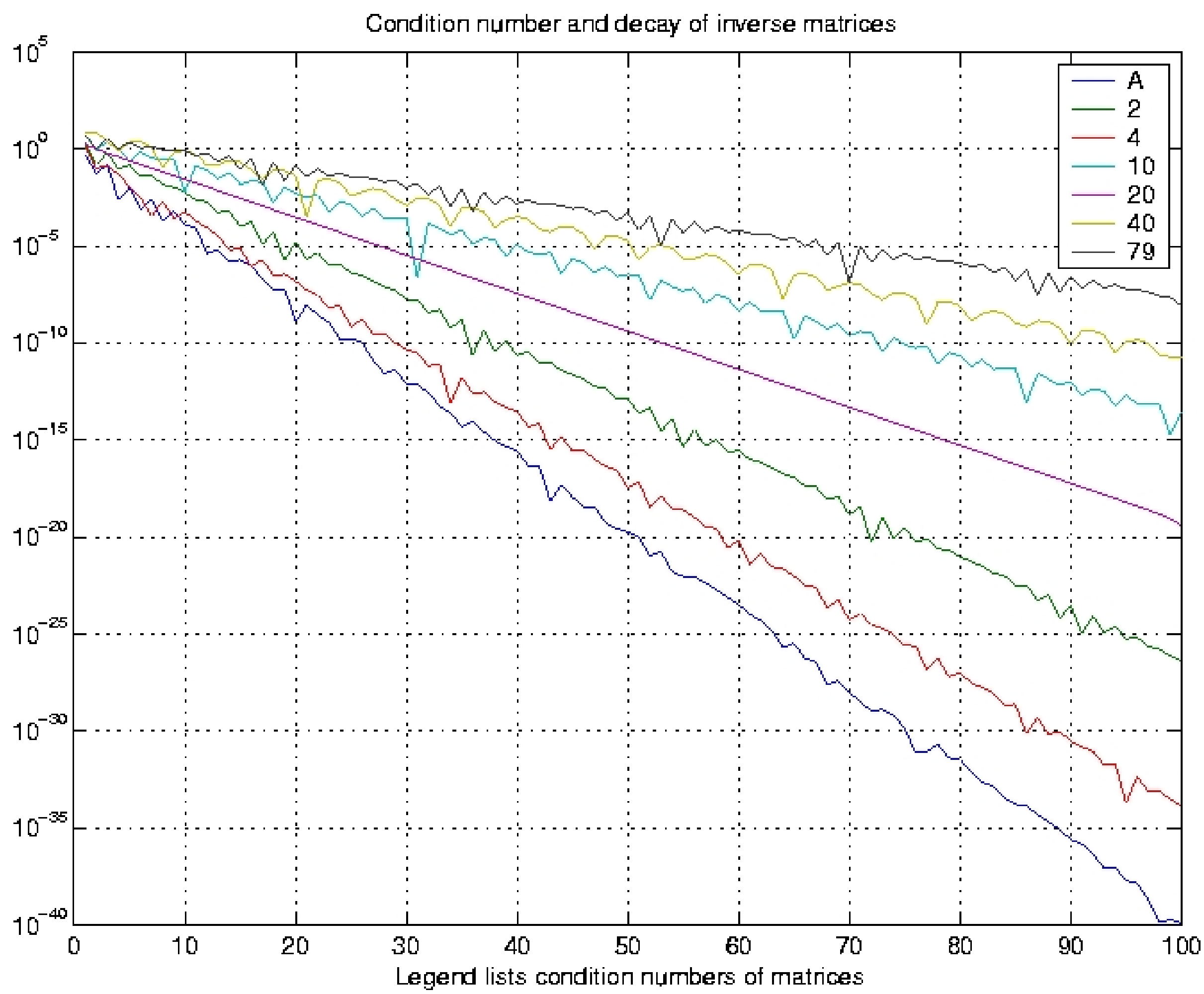
$$d_\alpha(x) = e^{-\alpha|x|} \text{or } d_\alpha(x) = (1+|x|)^{-\alpha}, \alpha > 0.$$

If A is invertible then

$$|(A^{-1})_{k,l}| \le \beta d_\gamma(|k-l|).$$

The parameters β and γ depend only on α and on cond(A) (also true for pseudo-inverse).

Thus: decay of off-diagonal entries of A^{-1} depends crucially on condition number of A!



Matrix off-diagonal decay and FIR-equalization

Basic setup: After A/D conversion at receiver, received sample at time index k is

$$y(k) = \sum_{n=1}^{L} h(L-n)x(k-L+n)$$

In matrix notation $y^{(k)} = Hx^{(k)} + \varepsilon$, where

$$H = \begin{bmatrix} \mathbf{h}_{L-1} & \mathbf{h}_{L-2} & \dots & \mathbf{h}_0 & \dots & 0 \\ 0 & \mathbf{h}_{L-1} & \mathbf{h}_{L-2} & \dots & \mathbf{h}_0 & 0 \\ \vdots & \ddots & & \ddots & & 0 \\ 0 & \dots & 0 & \mathbf{h}_{L-1} & \dots & \mathbf{h}_0 \end{bmatrix}$$

$x^{(k)} = [x(k-L+1), \dots, x(k+N-1)]^T$,
$y^{(k)} = [y(k), \dots, y(k+N-1)]^T$, ε is AWGN.
With M times oversampling $\mathbf{h}(k)$ becomes

$$\mathbf{h}_k = \begin{bmatrix} h(kT) \\ h(kT - \frac{T}{M}) \\ \vdots \\ h(kT - \frac{(M-1)T}{M}) \end{bmatrix}$$

H is $MN \times (N+L-1)$ matrix.

Matrix off-diagonal decay and equalization

Even if taps of channel H decay exponentially, large delay spread + deep fades will destroy off-diagonal decay of (left)-inverse H^+.

Any linear equalizer implicitly relies on H^+.

Well-known: With SIMO or oversampling at receiver we can construct FIR equalizers. For M-fold oversampling and channel with L taps with probability 1 there exists FIR equalizer with $L/(M-1)$ taps [Bresler '98].

MMSE-optimal equalizer

$$H_\sigma^+ := (H^*H + \sigma^2 I)^{-1} H^*$$

is not FIR, but due to oversampling H_σ^+ has non-trivial null-space which allows for FIR left-inverses H_{FIR}^+ which can be written as

$$H_{FIR}^+ = H_\sigma^+ + Null(H_\sigma^+)$$

Problems:

(i) taps of h are not completely random, therefore the "probability-0 case is not so unlikely".

(ii) H^{-1} and $(H^*H + \sigma^2 I)^{-1}H^*$ may have very slow decay. In this case we need to "dig deep" into nullspace of H^+ to construct FIR-filters. This leads to noise-enhancement and can increase condition number of equalizer: potentially severe loss of performance

(iii) Computing FIR filter with $L/(M-1)$ taps with standard methods is very expensive for large number of taps

Linear Precoding

Well-known idea: can move equalizer from receiver to transmitter: precoding/pre-equalizer.

Create virtual channel through precoding that allows for better and faster FIR equalizer design [Z. Ding, M. Cioffi, H. Sampath, A. Paulraj, A. Scaglione, G. Giannakis,...]

Assumptions:
Power constraint at transmitter is P=1
Channel impulse response h is normalized to 1
Channel H is frequency-selective, number of taps $\leq L$
Noise ε is AWGN with covariance matrix $\sigma^2 I$
Transmitter symbols are white

Precoding s.t. power constraint $\|Fx\| \le 1$:

$$x_{pre} = \frac{1}{\rho} Fx, \quad y = Hx_{pre} + \varepsilon, \qquad \tilde{x}_{pre} = \rho y$$

Expected mean-square error:

$$\mathcal{E}(\|x - \tilde{x}_{pre}\|^2) = \|HF - I\|_F^2 + \sigma^2 \rho^2$$

Optimal MMSE precoder F is given by

$$F = H^*(HH^* + \sigma^2 I)^{-1}, \qquad \text{with } \rho = \|F\|_F$$

Compare this to MMSE post-equalization:

$$y = Hx + \varepsilon, \quad \tilde{x}_{post} = Gy,$$

where

$$G = (HH^* + \sigma^2 I)^{-1} H^*$$

Mean-square error: $\mathcal{E}(\|x - \tilde{x}_{post}\|^2) =$

$$\|(H^*H + \sigma^2 I)^{-1} H^*H - I\|_F^2 + \sigma^2 \|(H^*H + \sigma^2 I)^{-1} H^*\|_F^2$$

Thus from MSE viewpoint: same performance

Advantage of precoding: burden of computational work is shifted to transmitter.

Stochastic error analysis

In practice CSI is less accurate at transmitter than at receiver. How does this affect pree-qualization? Consider zero-forcing precoder with power constraint.

Theorem: Assume estimated channel $\tilde{H}$ is

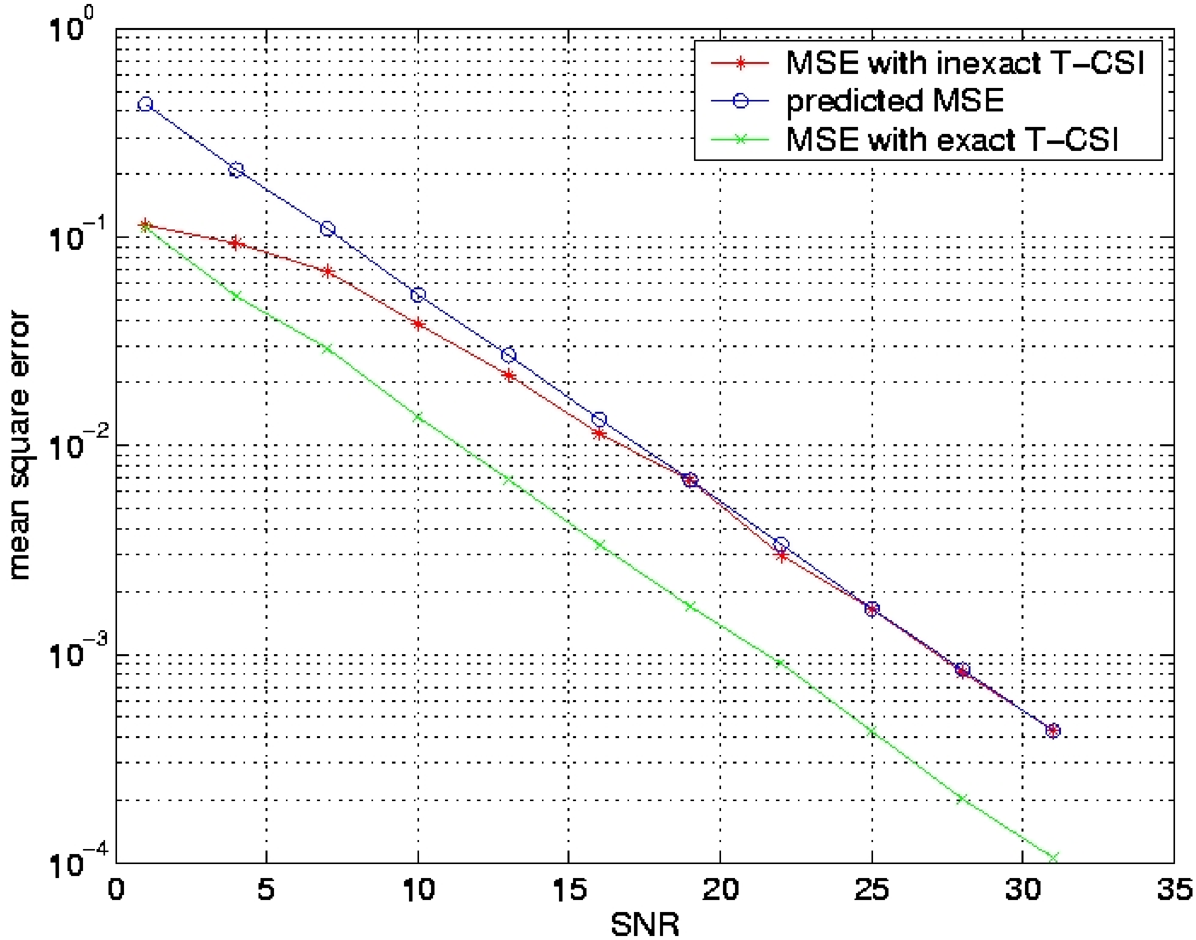
$$\tilde{H} = H + E$$

and E is AWGN with variance τ^2. Let $F = \tilde{H}^+ := \tilde{H}^*(\tilde{H}\tilde{H}^*)^{-1}$ and let $y = H(\rho F x) + \varepsilon$ where ε is AWGN with variance σ^2 and $\rho = 1/\|F\|_F$. Let the received signal be $\tilde{x} = \frac{1}{\rho} y$. If $\|H^*(HH^*)^{-1}\| \|E\| < 1$ and for τ not too large

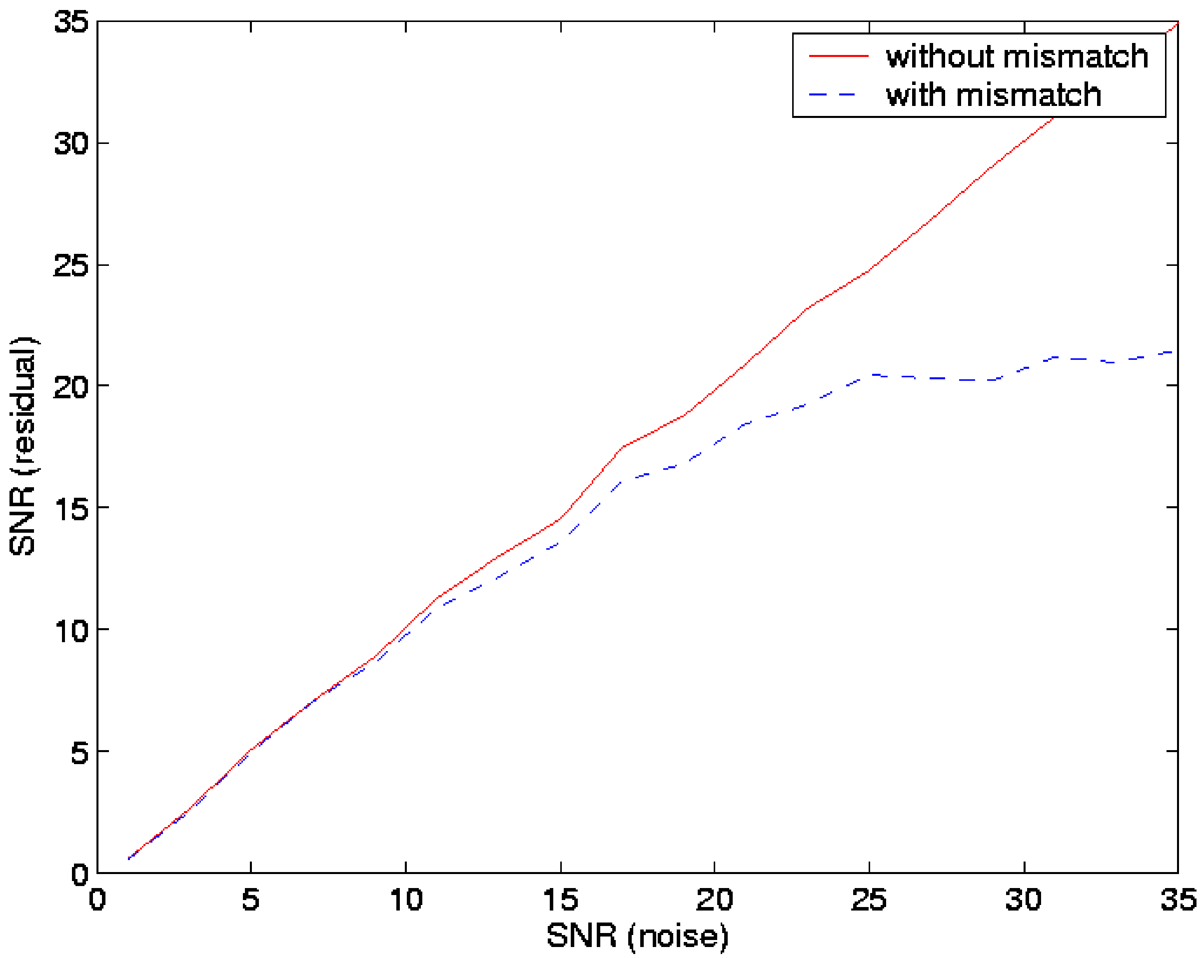
$$\mathcal{E}(\|x - \tilde{x}\|^2) = \|H^*(HH^*)^{-1}\|_F^2 \cdot (\tau^2 + \sigma^2)$$

Similar result holds for post ZF. Modification for MMSE instead of ZF straightforward.

If $\tau \ll \sigma$ we have essentially the same accuracy as with exact CSI. But if $\tau \approx \sigma$ theory predicts a 3dB loss compared to exact CSI.

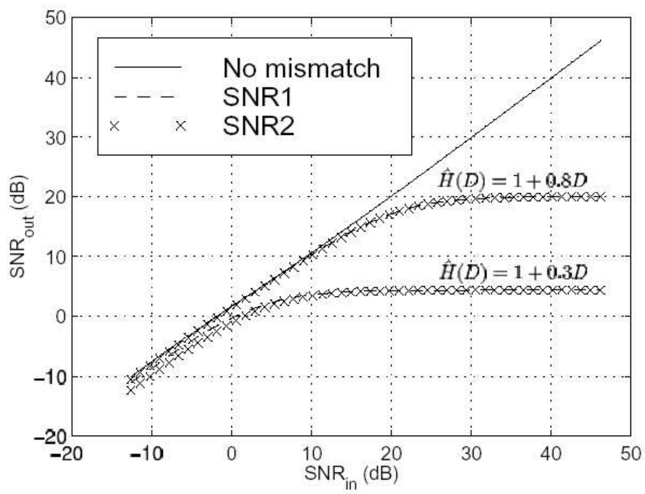


MSE with inexact T−CSI
predicted MSE
MSE with exact T−CSI
mean square error
SNR
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
0
5
10
15
20
25
30
35



without mismatch
with mismatch
SNR (residual)
SNR (noise)
0
5
10
15
20
25
30
35

DFE-FIR equalization with channel mismatch



Similar behavior reported for Tomlinson-Harashima precoding in case of channel mismatch at transmitter [Shi-Wesel, Proc. Asilomar 1998]

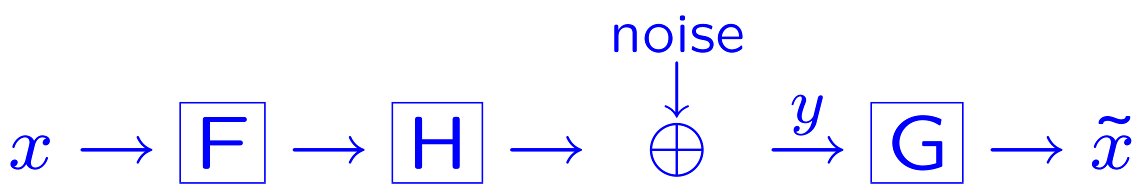
More research necessary

Precoding alone may not be sufficient.

Using only few largest taps for pre-equalization or post-equalization may lead to large channel mismatch, hence significant performance loss.

Joint equalization

Allow equalizer at transmitter and at receiver:



Optimal joint equalizer is derived by minimizing MSE(F,G). Explicit solution by [Sampath-Paulraj '99, Scaglione et al.,'02]: Let

$$V \Lambda V^* = \frac{1}{\sigma^2} H^* H,$$

thus V contains scaled right singular vectors of H. Then

$$G_{opt} = [(F_{opt})^*(F_{opt}) + \sigma^2 I]^{-1}(F_{opt})^*$$

$$F_{opt} = V\Phi$$

where Φ is a diagonal matrix that depends on singular values of H and on power P.

Our goal: construct (simple) precoder F such that HF has fast off-diagonal decay ($\approx$ few taps) and/or G can be constructed to have fast off-diagonal decay.

Does MSE-optimal joint equalizer allow this?

No! Answer lies in theory of matrices with off-diagonal decay.

If A has off-diagonal decay, then certain matrix factorizations such as LU and QR-decomposition inherit this decay, but other matrix factorizations do not such as SVD or eigenvalue decomposition! [T.S.,'04]

Simple example: Let A be biinifinite Toeplitz matrix. Eigenvectors of A are $e^{2\pi i\omega x}$, thus have no temporal decay at all even if A has exponential off-diagonal decay.

Hence $F_{opt} = V\Phi$ and G_{opt} will in general not have any off-diagonal decay and therefore MSE-optimal joint equalizer not useful for our purposes

Another approach to designing joint equalizer: Solve optimization problem:

$$\text{minimize } MSE(F, G)$$

subject to $\|F\|_F \leq 1$ and G has at most m taps

This leads to complicated non-convex global optimization problem.

Even if it can be solved, the solution is too complicated to compute for practical purposes.

Instead: find F such that HF has fast off-diagonal decay and cond(HF) is small.

Why? Because then noise will not be amplified much by G and we can truncate HF and/or G to simplify receiver.

Even if pre-equalizer F alone is not sufficient, right choice of F can reduce number of taps in virtual channel HF.

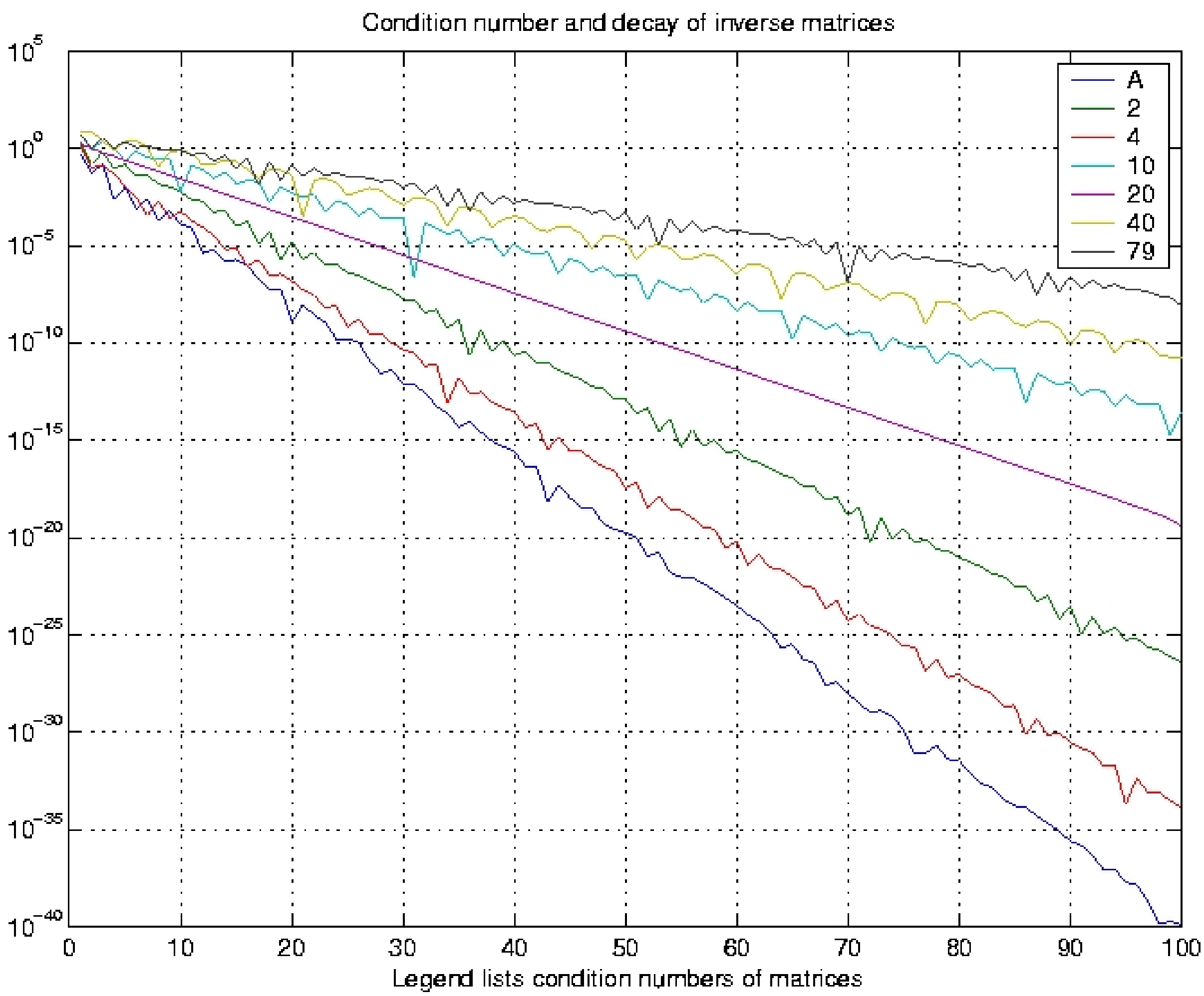
Good choice for F: MMSE-precoder.

$$F = H^*(HH^* + \sigma^2 I)^{-1}$$

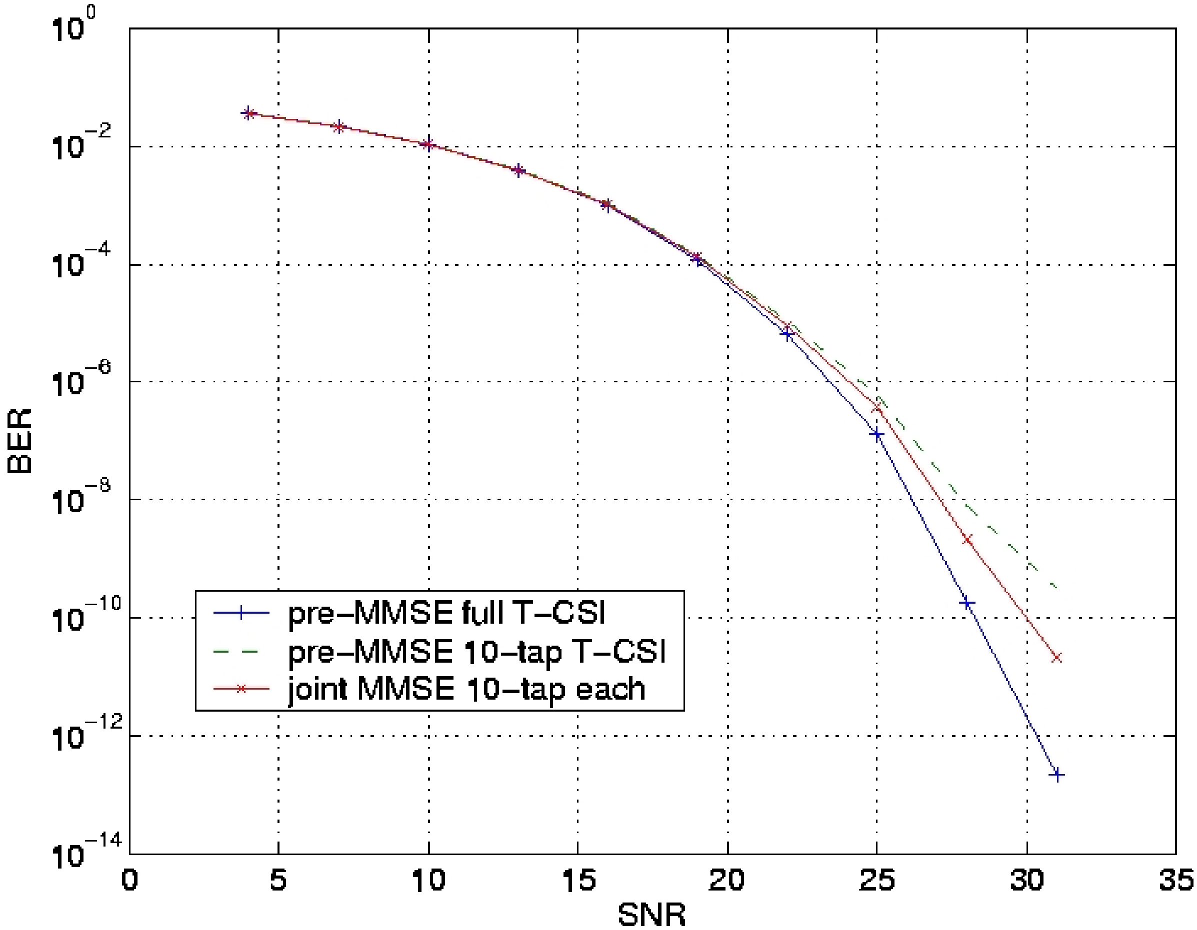
For given virtual channel HF optimal G is

$$G = [(FH)^*(FH) + \sigma^2 I]^{-1}(FH)^*$$

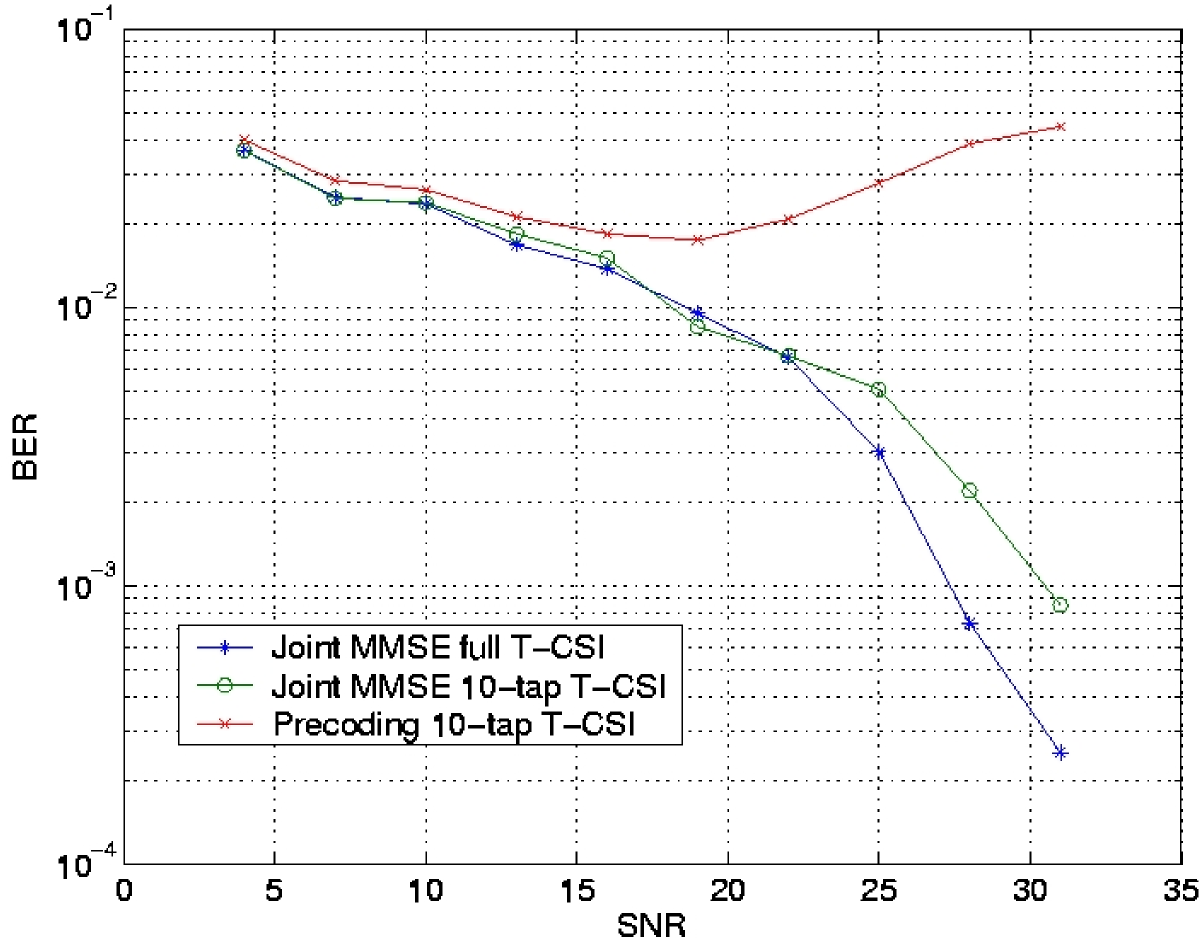
Other variations of the theme: could use only fixed number of taps at transmitter to construct approximate F. Or if HF has only few taps can use ML at receiver.



Condition number and decay of inverse matrices
10^5
10^0
10^−5
10^−10
10^−15
10^−20
10^−25
10^−30
10^−35
10^−40
0
10
20
30
40
50
60
70
80
90
100
Legend lists condition numbers of matrices
A
2
4
10
20
40
79



BER
SNR
10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
10^{-14}
0
5
10
15
20
25
30
35
pre–MMSE full T–CSI
pre–MMSE 10–tap T–CSI
joint MMSE 10–tap each



10^{-1}
10^{-2}
10^{-3}
10^{-4}
BER
0
5
10
15
20
25
30
35
SNR
Joint MMSE full T-CSI
Joint MMSE 10-tap T-CSI
Precoding 10-tap T-CSI

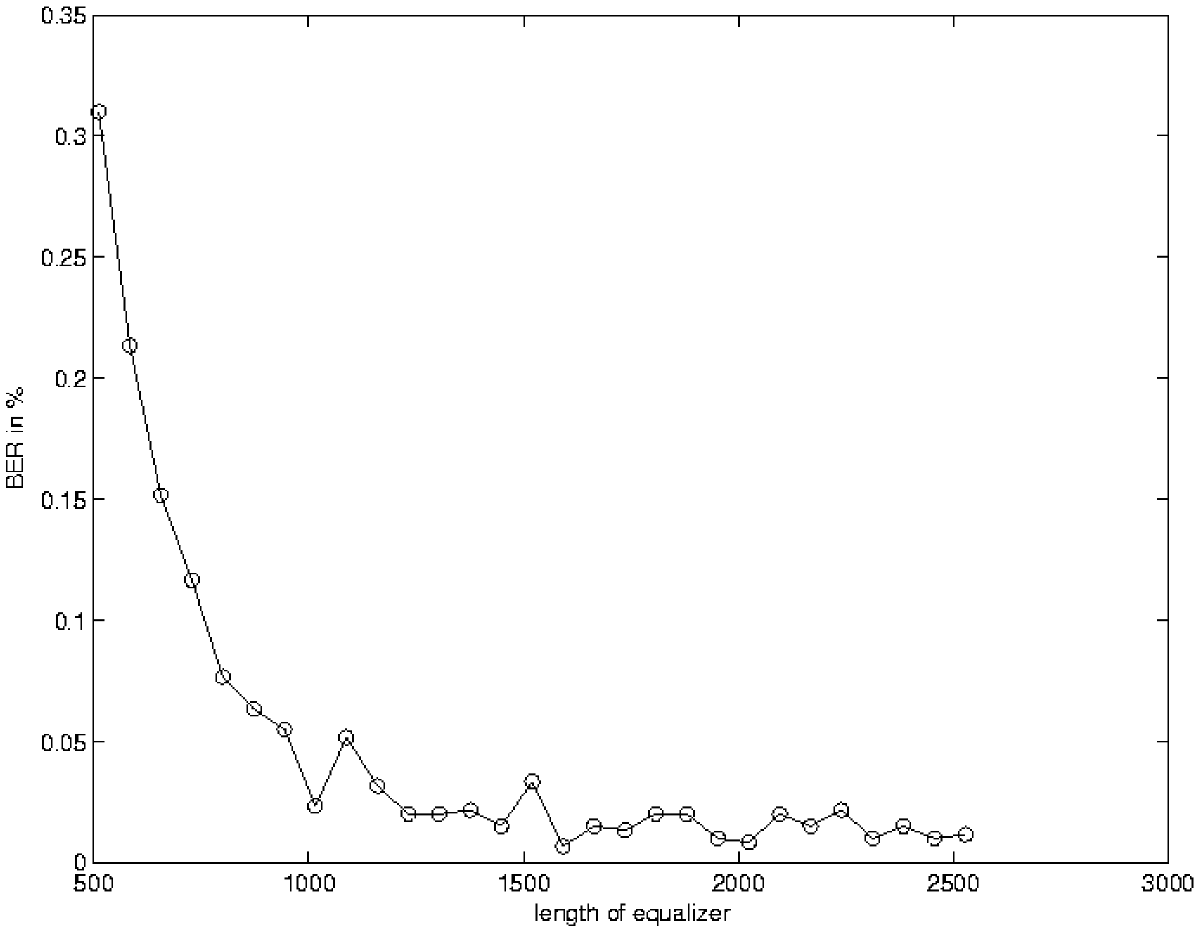
For noisy, large delay spread channels either transmitter or receiver has to do a thorough equalization.

Standard MMSE equalizers are computationally costly for large L. Need fast MMSE equalizers.

Fast FIR equalization via CG

Standard way of constructing FIR-equalizer for large number of taps is very costly. We have to compute FIR filter vector g (which generates FIR equalizer G), usually done via SVD.

May need large g to get good performance for FIR equalizer.



0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
BER in %
500
1000
1500
2000
2500
3000
length of equalizer

How to deal with huge computational costs for computing g?

Idea: Solve $Hg = e_k$ iteratively via conjugate gradient (CG) method. Here e_k is unit vector at index $k = (N + L - 1)/2$. Solution g is the FIR filter.

CG has been used before in equalization (CDMA,...), but in suboptimal way and its potential never fully explored.

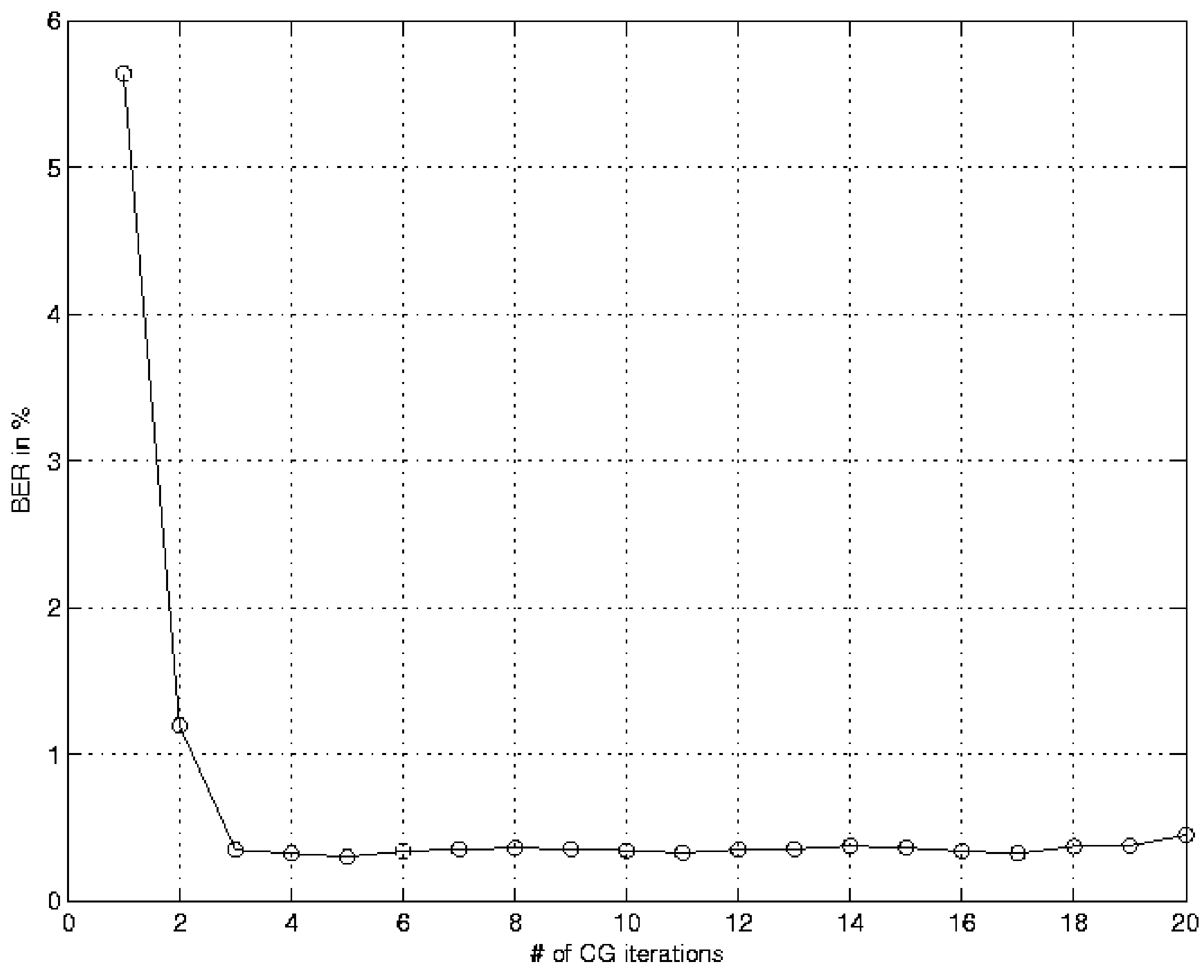
In theory CG converges to $g_{ZF} := H^{+}e_k$, which is the zero-forcing solution. But we want MMSE type solution.

We do not want to set up $(H^*H + \sigma^2 I)H^*$ explicitly, rather apply CG (or rather its variants CGLS, LSQR) directly to $Hg = e_k$.

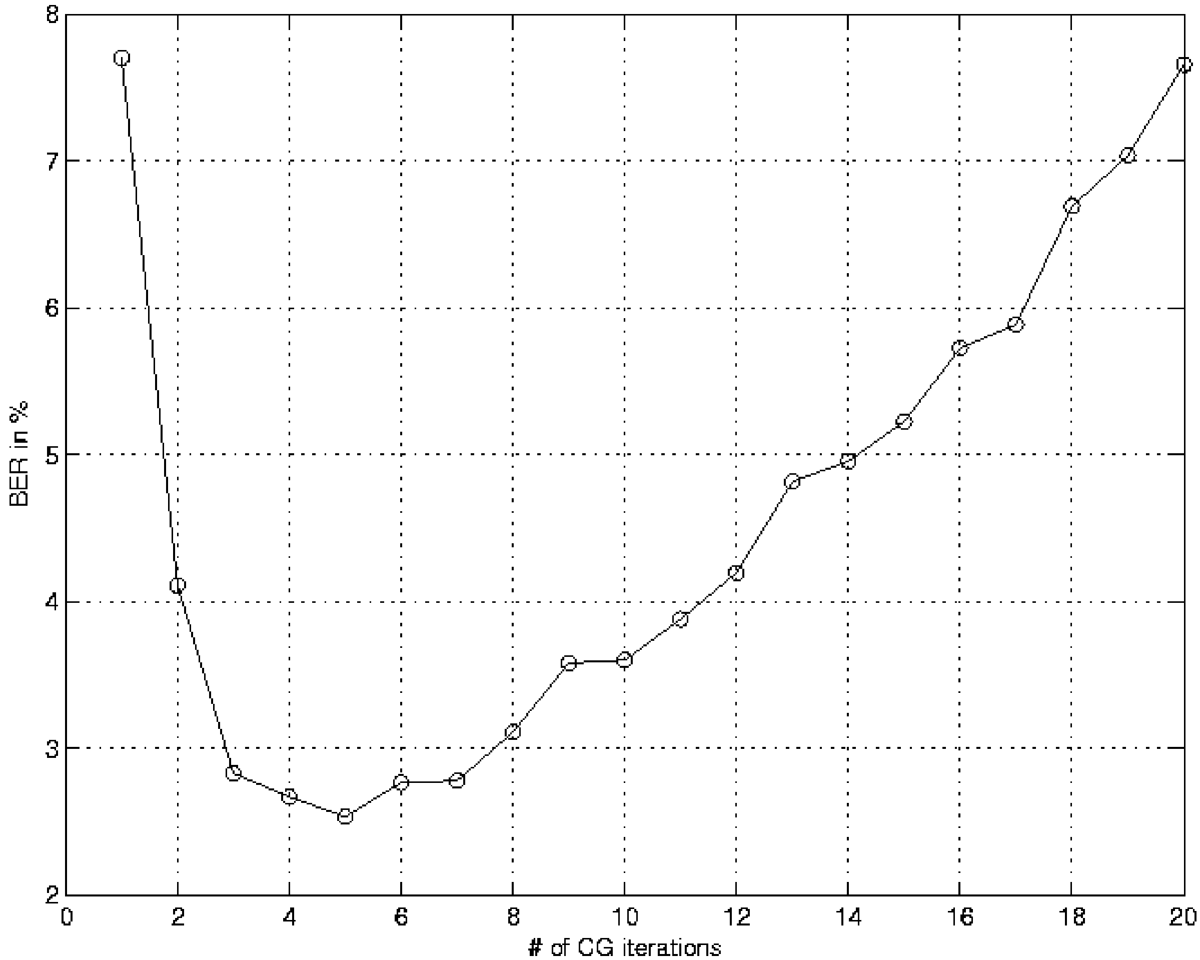
Important fact: CG is a regularization method, i.e., it converges first to direction of singular vectors associated to large singular values, and later to singular vectors of small singular values.

But: for noisy right-hand-sides monotone convergence may be lost

Thus: use number of iterations as stopping criterion to get MMSE-type solution

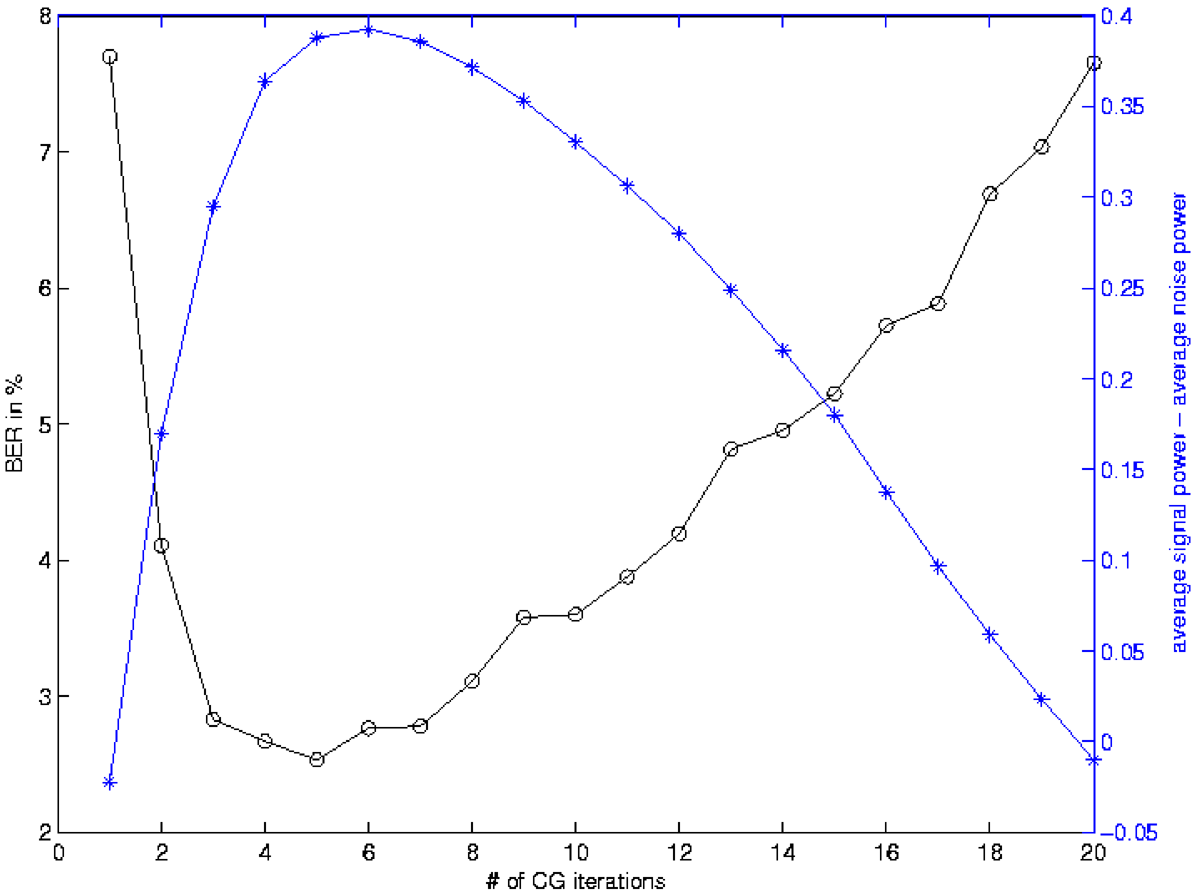


BER in %
# of CG iterations
0
1
2
3
4
5
6
0
2
4
6
8
10
12
14
16
18
20



Right choice of stopping criterion is delicate and crucial for success of CG!

We have found robust stopping criteria



Nice facts: H is block-Toeplitz matrix with vector-valued blocks. Matrix-vector multiplications Hx_n and $H^* z_n$ can be done by FFTs.

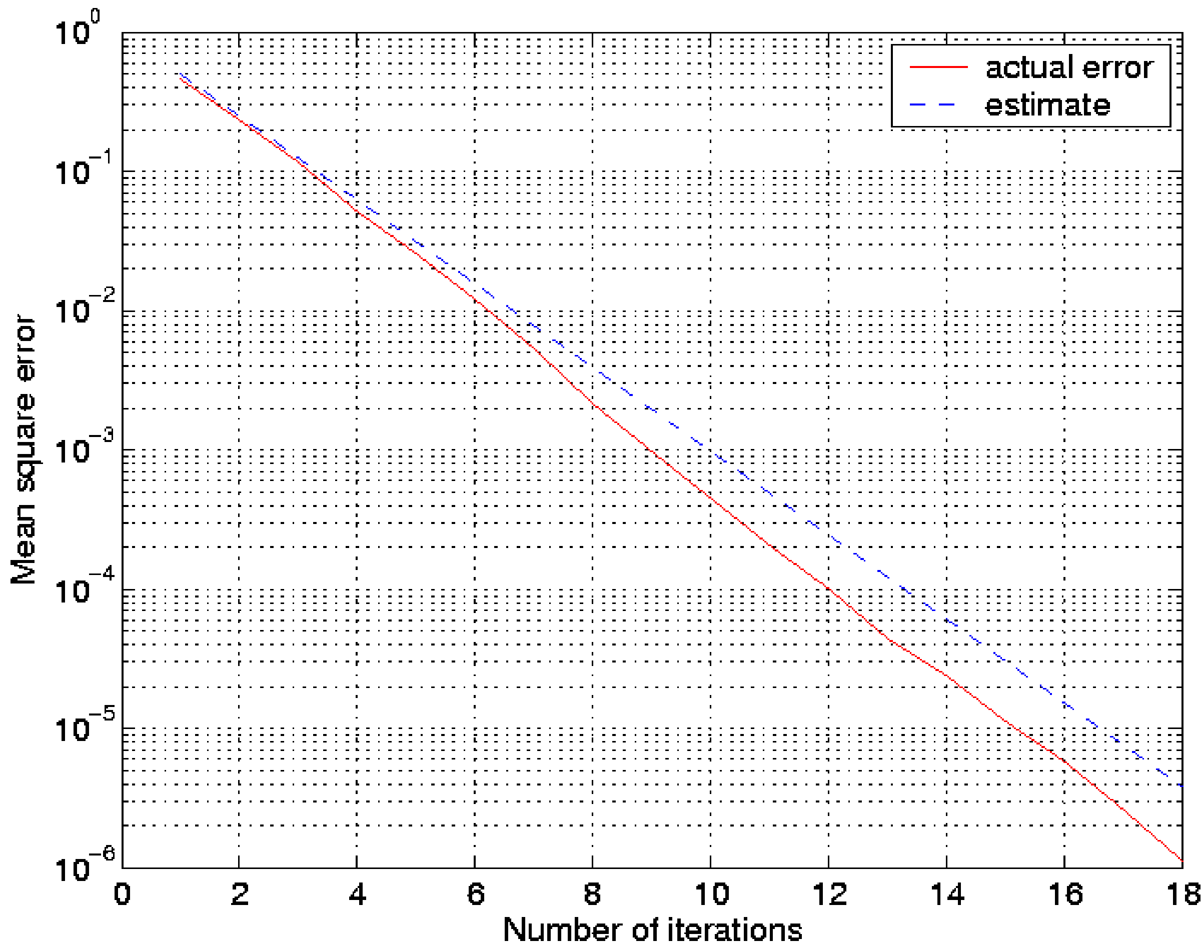
CG needs less iterations with increasing noise, often only 2-3 iterations are enough. Hence comp. costs are $\mathcal{O}(KM(N+L)\log(N+L))$, where K is often very small.

Additional preconditioning can be used.

Number of iterations

Theorem [T.S.2004]: Assume taps of impulse response are iid Gaussian. Assume we have either oversampling by factor a at receiver, or use a transmit antennas or use rate back-off by factor of a. Then expected rate of convergence of CG at k-th iteration is given by $(1/\sqrt{a})^k$.

Thus if $a = 4$ error after 10 iterations is 10^{-3}.



actual error
estimate
Mean square error
Number of iterations
10^0
10^-1
10^-2
10^-3
10^-4
10^-5
10^-6
0
2
4
6
8
10
12
14
16
18

Precoding = Preconditioning

Can exploit vast theory of preconditioners to construct precoders. Theory of preconditioner design is connected to matrix algebras with off-diagonal decay.

Extension to MIMO

Can use Gohberg-Semencul formula for inverse block Toeplitz matrix to accelerate computations further.

Theory of matrices with off-diagonal decay explains why truncated polyphase approach for computing FIR filter is very problematic for channels with deep fades.