Universal Decoding for Frequency-Selective Fading Channels

Ofer Shayevitz and Meir Feder

Universal Decoding- Background

Problem Setting:

- Unknown channel from a parametric family of channels
- Channel law - $p_\theta(\boldsymbol{y}|\boldsymbol{x})$, $\theta \in \Theta$ - unknown
- Codebook of M codewords $\mathcal{C} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$, $\boldsymbol{x}_i \in \mathcal{X}^n$
- A decoder $\Omega : \mathcal{Y}^n \mapsto \{1, \ldots, M\}$
- If θ known - Optimal ML decoder Ω^*:

$$i \Leftarrow \max_i p_\theta(\boldsymbol{y}|\boldsymbol{x}_i)$$

- Proposed decoders for the case of unknown θ:
  - Generalized Likelihood Ratio Test (GLRT) - Joint ML estimate of θ and the codewords

$$i \Leftarrow \max_i \sup_\theta p_\theta(\boldsymbol{y}|\boldsymbol{x}_i)$$

  - Training sequence for channel estimation, then ML decoding
- "Universal decoder": A decoder Ω that is independent of the channel parameters θ, yet performs (asymptotically) as well as the optimal ML decoder for that θ
- Both GLRT and training approach are not universal !

- *Competitive Minimax* criterion for universal decoding [*Feder,Merhav* 02'] :

$$\Omega_n \Leftarrow \inf \sup_{\theta \in \Theta} \frac{P_e(\Omega_n, \theta)}{P_e(\Omega^*, \theta)} \triangleq K_n$$

- The decoders are *universal in the error exponent sense* if K_n is subexponential in n

- Competitive Minimax w.r.t a fraction ξ of the ML exponent:

$$\Omega_n \Leftarrow \inf \sup_{\theta \in \Theta} \frac{P_e(\Omega_n, \theta)}{\left[P_e(\Omega^*, \theta)\right]^{\xi}} \triangleq K_n^{\xi}$$

OFDM Channel Model

- L bands and K time-points

$$Y = AX^{(i)} + Z, \qquad X^{(i)}, Y, Z \in \mathbb{R}^{L \times K}$$

- $X^{(i)}$ - the i-th codeword
- $A = diag\{\boldsymbol{a}\}$, $\boldsymbol{a}$ is a vector of band fading coefficients
- Z - matrix of i.i.d. normal rv's $\sim N(0,1)$
- Codewords are assumed not co-linear in any band
- Similar results are valid for complex channels

- A decoder for the OFDM channel :

$$\Omega : \mathbb{R}^{KL} \mapsto \{1, \ldots, M\}$$

- Define the decoder's *power error exponent* :

$$E^{\Omega}(\boldsymbol{a}) \triangleq \lim_{r \to \infty} -\frac{1}{r} \log P_e(\Omega, \sqrt{r}\boldsymbol{a})$$

- $E^{\Omega}(\boldsymbol{a})$ is the asymptotic slope of error probability as a function of the SNR on a logarithmic scale

- Let $\boldsymbol{x}_\ell^{(i)}$ be the ℓ-th row of the codeword matrix $X^{(i)}$. Define

$$P_\ell^{(i)} = \| \boldsymbol{x}_\ell^{(i)} \|^2 , \qquad \rho_\ell^{(i,j)} = \frac{\langle \boldsymbol{x}_\ell^{(i)}, \boldsymbol{x}_\ell^{(j)} \rangle}{\sqrt{P_\ell^{(i)} P_\ell^{(j)}}}$$

- The ML power error exponent :

$$E^*(\boldsymbol{a}) = \frac{1}{8} \min_{i \neq j} \left\{ \sum_{\ell=0}^{L-1} \left(P_\ell^{(i)} + P_\ell^{(j)} - 2\sqrt{P_\ell^{(i)} P_\ell^{(j)}} \rho_\ell^{(ij)} \right) a_\ell^2 \right\}$$

The Minimax Criterion Modified

- Modifications: SNR-asymptotic, Family of decoders
- Let $\mathcal{C}$ be a given codebook, $\mathcal{F}$ a family of decoders
- We seek $\Omega \in \mathcal{F}$ achieving a maximal fraction $\xi = \xi^*$ of the ML exponent uniformly over all channels $\boldsymbol{a}$

$$\inf_{\Omega \in \mathcal{F}} \sup_{\boldsymbol{a}} \frac{P_e(\Omega, \boldsymbol{a} \,|\, \mathcal{C})}{[P_e(\Omega^*, \boldsymbol{a}, \,|\, \mathcal{C})]^{\xi}} \approx \inf_{\Omega \in \mathcal{F}} \sup_{\substack{\|\boldsymbol{a}\|=1 \\ r \gg 0}} \left(\xi E^*(\boldsymbol{a}) - E^{\Omega}(\boldsymbol{a})\right) r$$

$$\Rightarrow \quad \xi^* = \sup_{\Omega \in \mathcal{F}} \inf_{\|\boldsymbol{a}\|=1} \frac{E^{\Omega}(\boldsymbol{a})}{E^*(\boldsymbol{a})}$$

Pairwise Decoders

- Any Decoder Ω can be decomposed into *pairwise decoders*

$$\Omega^{ij} : \mathbb{R}^{KL} \mapsto \{i, j\}$$

- A set of Ω^{ij} plus inconsistency resolving rule, define Ω

- The *pairwise power error exponent* is defined as

$$E_{ij}^{\Omega}(\boldsymbol{a}) \triangleq \lim_{r\to\infty} -\frac{1}{r} \log P_e^{i\to j}(\Omega^{ij}, \sqrt{r}\boldsymbol{a})$$

- The pairwise minimal distance $d_{ij}(\Omega^{ij}\,; \boldsymbol{a})$ of codeword i to the separating surface of Ω^{ij} is related to $E_{ij}^{\Omega}(\boldsymbol{a})$ by

$$E_{ij}^{\Omega}(\boldsymbol{a}) = \frac{1}{2} d_{ij}^{2}(\Omega^{ij}\,; \boldsymbol{a})$$

- The decoder's exponent is dominated by the worst pair

$$E^{\Omega}(\boldsymbol{a}) = \min_{i \neq j} E_{ij}^{\Omega}(\boldsymbol{a})$$

Quadratic Decoders

- Ω is called a *quadratic decoder* if it can be decomposed

$$\Omega^{ij}(\boldsymbol{y}) = \begin{cases} i & \text{if} \quad \boldsymbol{y}^T H_{ij} \boldsymbol{y} \geq 0 \\ j & \text{otherwise} \end{cases}$$

- The GLRT is a quadratic decoder for the OFDM channel
- If $\mathcal{F}$ is a family of quadratic decoders, then the decoder achieving ξ^* is called a *Quadratic Minimax (QMM)* Decoder

Specifying the Family of Quadratic Decoders

- Considering only projections per band is sufficient
- The corresponding pairwise decoders are each dependent on a selection of a 2×2 symmetric matrix per band
- We restrict these matrices to be diagonal
- We focus on the family $\mathcal{F}$ of all quadratic decoders with such pairwise components
- The GLRT is a member of the family $\mathcal{F}$

Bounding the Power Error Exponent

- For $\Omega \in \mathcal{F}$, $E^{\Omega}(\boldsymbol{a})$ is a solution of an optimization problem
- This solution has no analytical expression in general
- Solving a modified problem, a lower bound $\hat{E}^{\Omega}(\boldsymbol{a})$ is found
- To maximize $\hat{E}^{\Omega}(\boldsymbol{a})$ only a family $\mathcal{F}^* \subset \mathcal{F}$ needs to be considered
- Decoders $\Omega \in \mathcal{F}^*$ have pairwise components each dependent on a **single** parameter λ_{ij}

- For any selection of λ_{ij}, the lower bound is given by

$$\hat{E}^{\Omega}(\boldsymbol{a}) = \frac{1}{2} \min_{i \neq j} \left\{ \frac{1}{1+\lambda_{ij}} \sum_{\ell=0}^{L-1} P_{\ell}^{(i)} (1 - |\rho_{\ell}^{(ij)}|) a_{\ell}^{2} \right\}$$

- We use $\hat{E}^{\Omega}(\boldsymbol{a})$ instead of $E^{\Omega}(\boldsymbol{a})$

- The corresponding lower bound $\hat{\xi}^*$ for ξ^* is

$$\xi^* \geq \sup_{\Omega \in \mathcal{F}^*} \inf_{\|\boldsymbol{a}\|=1} \frac{\hat{E}^{\Omega}(\boldsymbol{a})}{E^*(\boldsymbol{a})} \triangleq \hat{\xi}^*$$

- We seek $\Omega \in \mathcal{F}^*$ attaining the bound $\hat{\xi}^*$ above

Optimal Weights Selection

- A decoder $\Omega \in \mathcal{F}^*$ is defined via its weights λ_{ij}
- For any selection of λ_{ij}, *Critical channels* are those attaining the infimum

$$\inf_{\|\boldsymbol{a}\|=1} \frac{\hat{E}^{\Omega}(\boldsymbol{a})}{E^*(\boldsymbol{a})}$$

- In general, the critical channels depend on λ_{ij} and on the codebook, and are hard to determine

However...

- Both $E^*(\boldsymbol{a})$ and $\hat{E}^{\Omega}(\boldsymbol{a})$ are linear in a_l^2
- For $\|\boldsymbol{a}\| = 1$: Let $\boldsymbol{b} = (a_0^2, a_1^2, \ldots, a_{L-2}^2)$
- $E^*(\boldsymbol{b})$ and $\hat{E}^{\Omega}(\boldsymbol{b})$ are both polyhedral concave functions defined over the simplex

$$S = \left\{ \boldsymbol{b} \,\middle|\, b_\ell \geq 0\,, \sum b_\ell \leq 1 \right\}$$

- Let $\mathcal{P}$ be the set of all extreme points of $E^*(\boldsymbol{b})$ over S

Theorem *The set $\mathcal{P}$ is finite and contains at least one critical channel for any selection of the weights λ_{ij}*

- $\mathcal{P}$ can be determined a-priori
- Let $\mathcal{P} = \left\{ \boldsymbol{b}^1, \ldots, \boldsymbol{b}^N \right\}$
- Let $\{\boldsymbol{a}^1, \ldots, \boldsymbol{a}^N\}$ be the corresponding set of channels
- The weights λ_{ij}^* attaining the lower bound $\hat{\xi}^*$ are

$$\lambda_{ij}^* = \frac{s^{(ij)}}{s^{(ji)}}, \quad s^{(ij)} = \min_n \left\{ \frac{\sum_\ell P_\ell^{(i)} (1 - |\rho_\ell^{(ij)}|)(a_\ell^n)^2}{E^*(\boldsymbol{a}^n)} \right\}$$

QMM Decoding Procedure

- Preprocessing - determine λ^*_{ij} by critical channel analysis
- Let $\boldsymbol{r}_\ell$ be the projection of the observation onto the subspace generated by the two codewords in each band
- Express each projection by $\boldsymbol{r}_\ell = \alpha_\ell \boldsymbol{x}_\ell^{(i)} + \beta_\ell \boldsymbol{x}_\ell^{(j)}$
- The QMM pairwise decoder prefers codeword i over j if :

$$\sum_\ell \left[P_\ell^{(i)} (1 - |\rho_\ell^{(ij)}|) \right] \alpha_\ell^2 \ \geq \ \lambda^*_{ij} \sum_\ell \left[P_\ell^{(j)} (1 - |\rho_\ell^{(ij)}|) \right] \beta_\ell^2$$

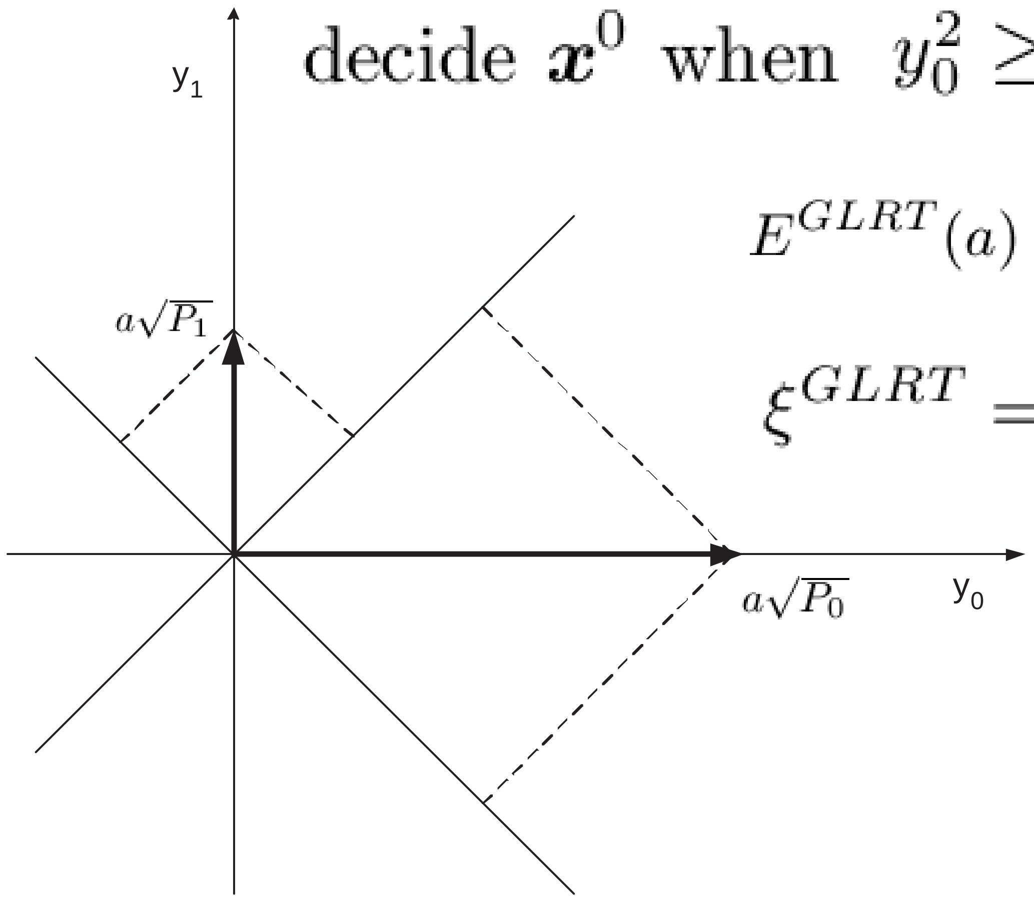
Example - Flat Fading Channel

- Flat fading channel (a - unknown)

$$y = ax + n\,, \quad n \sim N(0,1)$$

- Assume two codewords $\boldsymbol{x}^0 = (\sqrt{P_0}, 0)$ and $\boldsymbol{x}^1 = (0, \sqrt{P_1})$
- The critical channel here is $a = 1$
- For $\lambda = 1$ we get the GLRT. The QMM weight is

$$\lambda^* = \frac{P_0}{P_1}\,, \quad s^{(01)} = \frac{P_0}{P_0 + P_1}\,, \quad s^{(10)} = \frac{P_1}{P_0 + P_1}$$

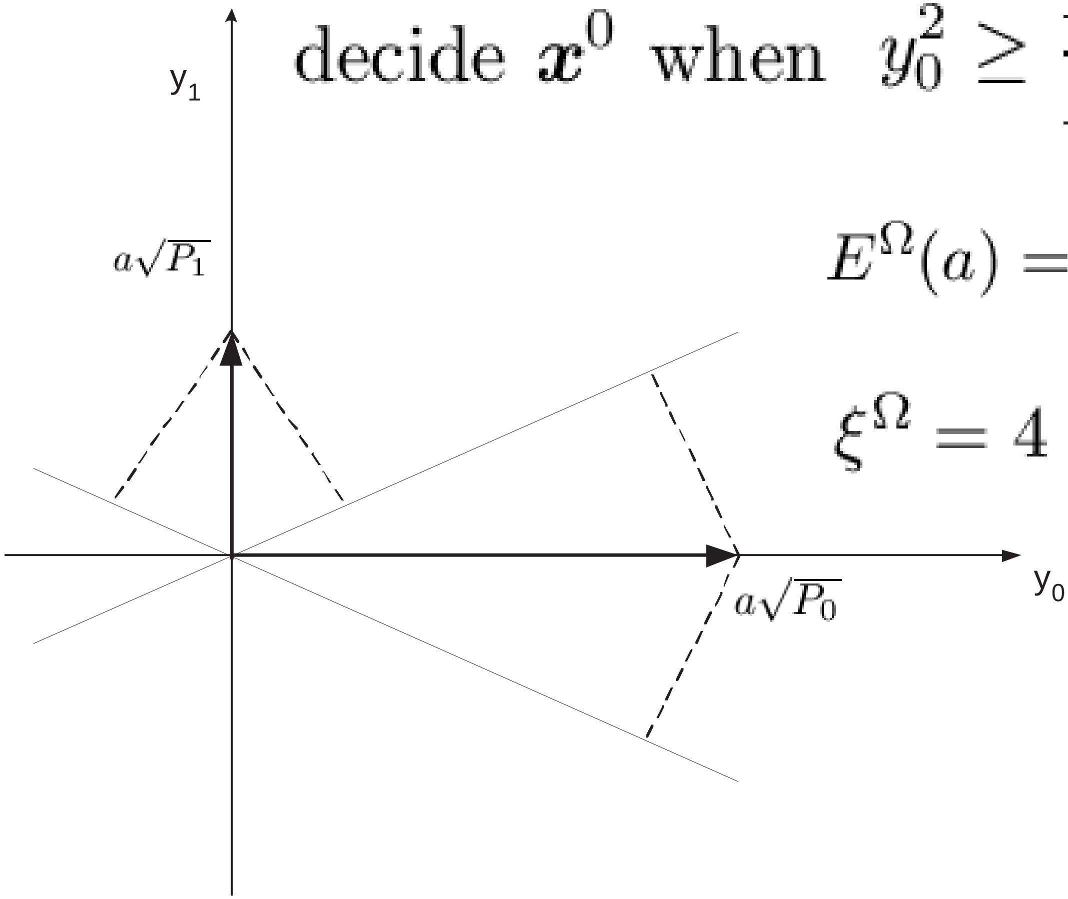


decide $\boldsymbol{x}^0$ when $y_0^2 \geq y_1^2$

$$E^{GLRT}(a) = \frac{1}{4}\min(P_0, P_1)a^2$$

$$\xi^{GLRT} = 2\,\frac{\min(P_0, P_1)}{P_0 + P_1}$$

GLRT for fading channel



decide $\boldsymbol{x}^0$ when $y_0^2 \geq \frac{P_0}{P_1} y_1^2$

$$E^{\Omega}(a) = \frac{1}{2}\frac{P_0 P_1}{P_0 + P_1} a^2$$

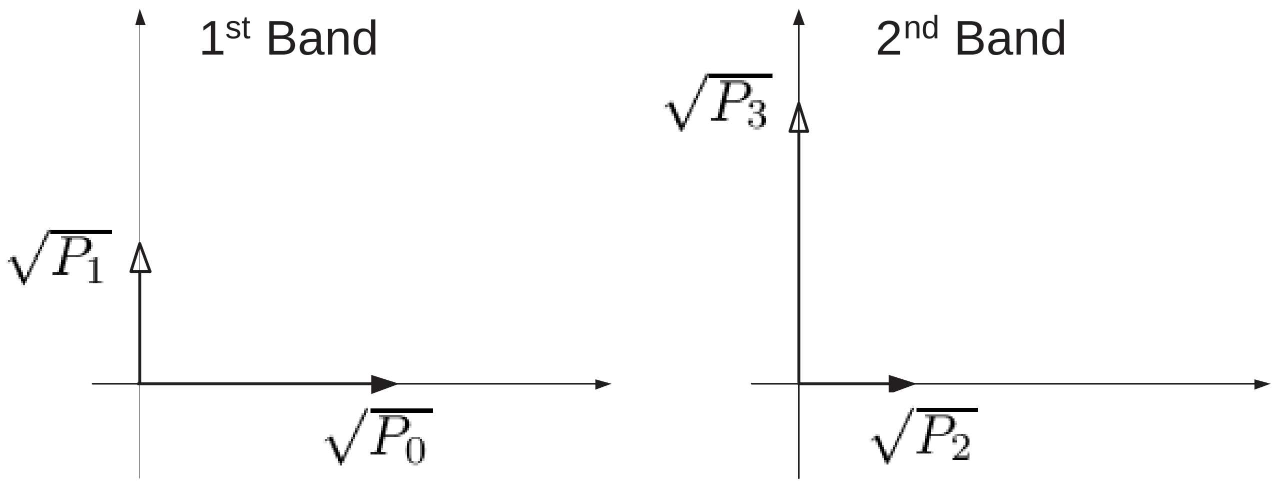
$$\xi^{\Omega} = 4\,\frac{P_0 P_1}{(P_0 + P_1)^2}$$

QMM for fading channel (attaining optimal ξ^*)

Example - 2×2 OFDM Channel

- Consider a 2×2 OFDM channel with two codewords

$$X^{(0)} = \begin{bmatrix} \sqrt{P_0} & 0 \\ \sqrt{P_2} & 0 \end{bmatrix}, \quad X^{(1)} = \begin{bmatrix} 0 & \sqrt{P_1} \\ 0 & \sqrt{P_3} \end{bmatrix}$$



- The simplex is the segment $[0,1]$, and the critical channels are the its vertices. The QMM weight is

$$s^{(01)} = \min\left\{\frac{P_0}{P_0+P_1}, \frac{P_2}{P_2+P_3}\right\}, \; s^{(10)} = \min\left\{\frac{P_1}{P_0+P_1}, \frac{P_3}{P_2+P_3}\right\}$$

$$\lambda^* = \frac{s^{(01)}}{s^{(10)}} \neq \lambda^{GLRT} = 1$$

- Even when the total power of both codewords is equal, the GLRT is generally different from the QMM

- The GLRT coincides with the QMM if and only if the codewords have equal power in both bands

QMM Versus GLRT

The GLRT's power error exponent can be upper bounded by

$$E^{GLRT}(\boldsymbol{a}) \leq \frac{1}{4} \min_{i \neq j} \left\{ \sum_{\ell=0}^{L-1} P_\ell^{(i)} (1 - |\rho_\ell^{(ij)}|) a_\ell^2 \right\}$$

Theorem *The proposed (suboptimal) QMM decoder always attains a guaranteed fraction* ξ^* *of the ML exponent that is equal or higher than that of the GLRT*

Utilizing Fading Relations

- Information about fading interdependency may be available
- Suppose the fading vector $\boldsymbol{a}$ is known to belong to a set A
- Assume A is scale invariant, and define

$$S = \left\{ \boldsymbol{b} \,\middle|\, \boldsymbol{b} = (a_0^2, \ldots, a_{L-2}^2)\,,\, \|\boldsymbol{a}\| = 1\,,\, \boldsymbol{a} \in A \right\}$$

- If S is polyhedral, it may be used instead of the simplex to find the QMM weights maximizing the lower bound over A

Example - Channel with Unknown Gain

- Assume that channel is known to be $\boldsymbol{a}^0$ up to a gain factor:

$$A = \{\boldsymbol{a} \,|\, \boldsymbol{a} = r\boldsymbol{a}^0 \,,\, r \in \mathbb{R}\}$$

- The set A is scale invariant by definition

- The corresponding set S is

$$S = \{\boldsymbol{b}^0\}\,, \quad \boldsymbol{b}^0 = ((a_0^0)^2, (a_1^0)^2, \ldots, (a_{L-2}^0)^2)$$

- The set S is polyhedral

- Obviously, $\boldsymbol{b}^0$ is the only extreme point of $E^*(\boldsymbol{b})$ over S

- The optimal weights are given by

$$s^{(ij)} = \frac{\sum_\ell P_\ell^{(i)}(1-|\rho_\ell^{(ij)}|)(a_\ell^0)^2}{E^*(\boldsymbol{a}^0)}$$

$$\lambda_{ij}^* = \frac{s^{(ij)}}{s^{(ji)}} = \frac{\sum P_\ell^{(i)}(1-|\rho_\ell^{(ij)}|)(a_\ell^0)^2}{\sum P_\ell^{(j)}(1-|\rho_\ell^{(ij)}|)(a_\ell^0)^2}$$

- This setting is equivalent to a flat fading channel (L=1)

Example - Multipath Channel

- Consider a discrete time channel with an impulse response

$$h[n] = c_0\delta[n] + c_1\delta[n-m]$$

- Path gains $c0, c1 \in \mathbb{R}$ unknown, path delay $0 < m < N$
- Convert the channel to a L-band OFDM channel $(L > N)$

$$a_\ell = \sum_{n=0}^{L-1} h[n] \exp\{-\frac{j2\pi n\ell}{L}\} = c_0 + c_1 \exp\{-\frac{j2\pi m\ell}{L}\}$$

- Defining $\alpha_\ell^m = \cos\{\frac{2\pi m\ell}{L}\}$ plus some algebra, we get

$$\Upsilon_{k\ell}^L \leq \frac{|a_\ell|^2}{|a_k|^2} \leq \Upsilon_{k\ell}^U$$

where

$$\Upsilon_{k\ell}^L \triangleq \min_{0<m<N} \frac{1-\alpha_\ell^m\,\alpha_k^m - |\alpha_\ell^m - \alpha_k^m|}{1-(\alpha_k^m)^2}$$

$$\Upsilon_{k\ell}^U \triangleq \max_{0<m<N} \frac{1-\alpha_\ell^m\,\alpha_k^m + |\alpha_\ell^m - \alpha_k^m|}{1-(\alpha_k^m)^2}$$

- Let the constraint set A be the set of all fading vectors $\boldsymbol{a}$ satisfying the above requirement

- The set A is scale invariant
- The set corresponding set S is the set of vectors satisfying

$$b_\ell - \Upsilon_{k\ell}^{U}\, b_k \leq 0\,, \quad \Upsilon_{k\ell}^{L}\, b_k - b_\ell \leq 0\,, \quad b_\ell \geq 0$$

- S is a polyhedra
- The set of extreme points of $E^*(\boldsymbol{b})$ over S contain at least one critical channel for any selection of λ_{ij}

Simulation Results

- Code : Based on Complex Field Coding (CFC) [*Wang, Giannakis* 03'] with and without training
- CFC helps to prevent co-linearity of codewords in the bands
- Training can be used with any code, since it prevents co-linearity inherently
- QMM is compared to the GLRT and to the training approach in terms of the attained ξ over random channels

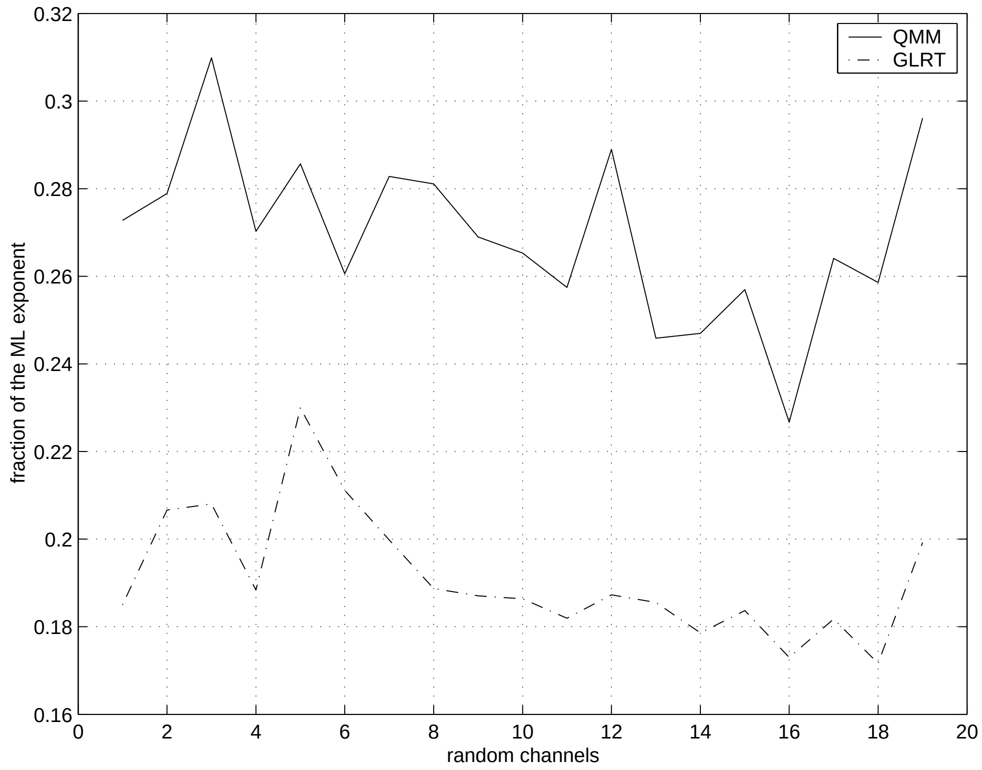


Figure 1: QMM vs. GLRT, 120 codewords CFC code, 3×4 channel

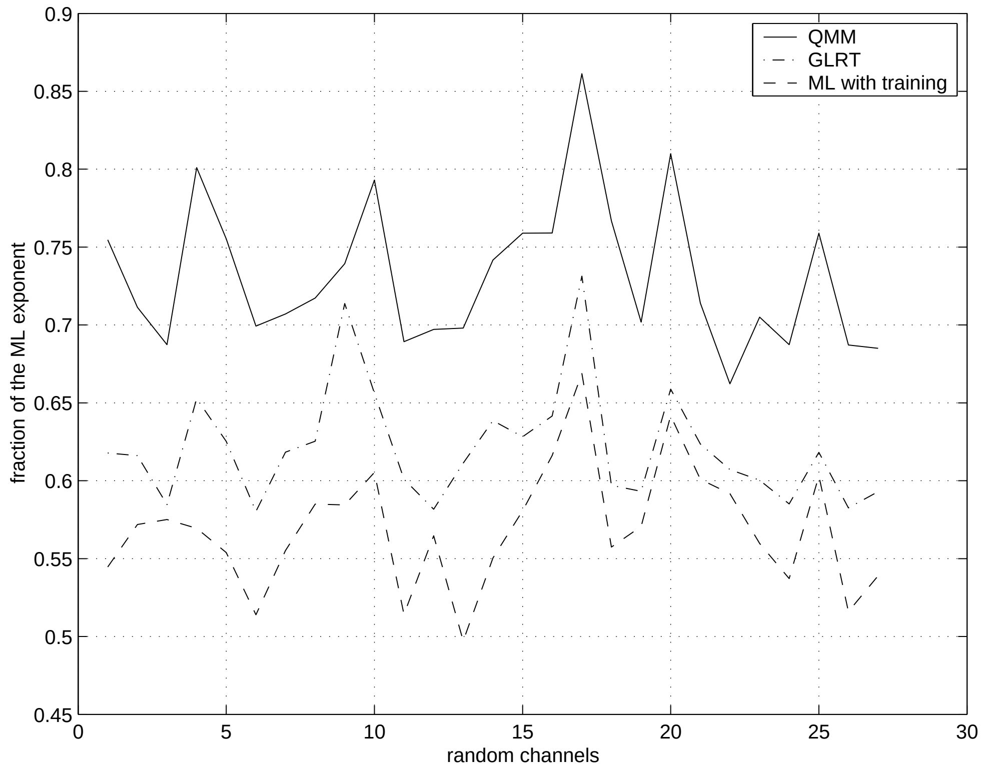


Figure 2: QMM vs. GLRT and ML-training, 28 codewords CFC code 3×3 channel with 1 training symbol per block

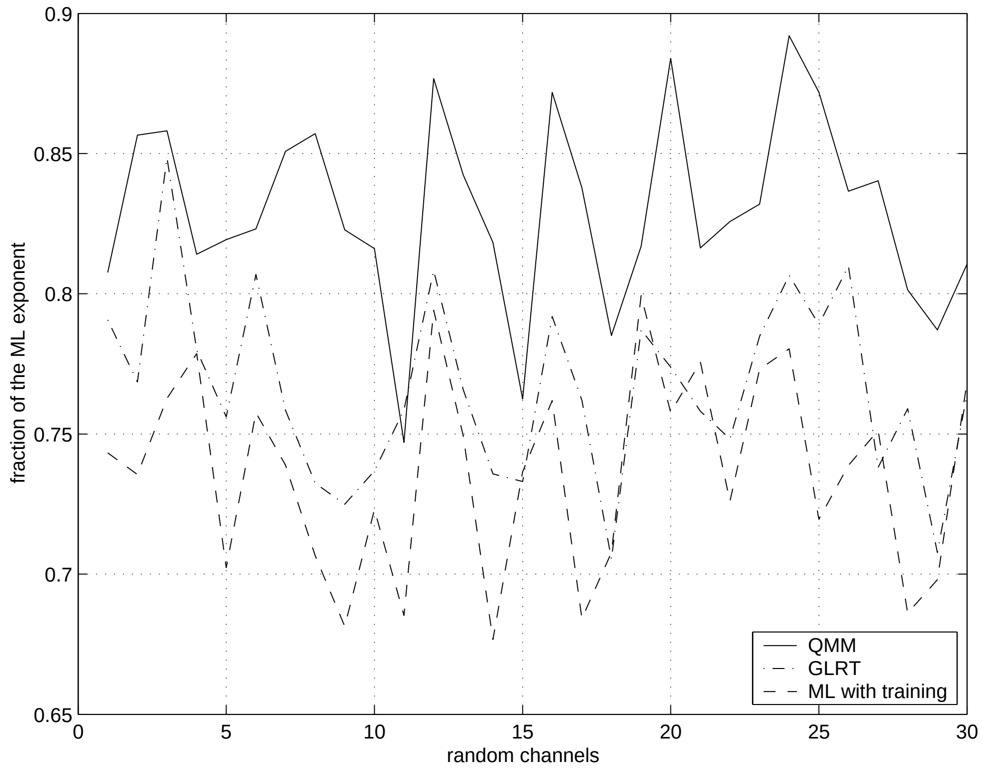


Figure 3: QMM vs. GLRT and ML-training, 28 codewords CFC code 3×3 channel with 2 training symbol per block

Summary

- Universal decoding for the OFDM setting was discussed
- A SNR-asymptotic minimax criterion was suggested
- The QMM decoder was constructively introduced
- The QMM complexity is comparable to that of the GLRT
- The QMM outperforms the GLRT in the minimax criterion
- Means of utilizing fading interdependency were suggested

Future Research

- Efficient QMM decoding methods should be explored
- How to design good codes for which the QMM has good performance ?
- Is there a "large" set of channels over which the resulting QMM decoder attains $\xi^* \to 1$?
- QMM performance under standard fading models (Rayleigh, Rician, etc.) should be investigated
- Extensions to unknown MIMO channels are being explored