Random Matrix Theory in Communications Engineering

Prof. Dr.-Ing. Ralf R. Müller

mueller@lnt.de

August 8, 2013

Chapter 1:

Multiuser Detection

The Multiple–Access Scheme

Definition 1 *A multiple–access scheme is an algorithm that defines how to generate the signals $x_1(t), x_2(t), \ldots, x_K(t)$ corresponding to user 1 to K from the discrete–time data streams $b_1[\mu], b_2[\mu], \ldots, b_K[\mu] \in \mathcal{A}$.*

In general, $x_k(t)$ depends on the data streams of all users.

In practice, $x_k(t)$ often depends on the data stream of user k only.

The symbol alphabet $\mathcal{A}$ is determined by the modulation scheme that is used, e.g. $\mathcal{A} = \{+1, -1\}$ for binary phase shift keying.

Linear Multiple–Access

Definition 2 *A multiple–access scheme is called linear if and only if the signal $x_k(t)$ is a linear combination of the data stream $b_k[\mu]$ for all users $k = 1 \ldots K$.*

This means

$$x_k(t) = \sum_{\mu=-\infty}^{+\infty} g_{k,\mu}(t) * b_k[\mu]\, \delta(t - \mu T_{\rm s})$$

for some symbol waveforms $g_{k,\mu}(t)$ with $T_{\rm s}$ denoting the symbol clock cycle.

In practice, symbol waveforms are often **in**variant to discrete time, i.e.

$$g_{k,\mu}(t) = g_k(t) \quad \forall\, \mu.$$

The Chips

In practice, the symbol waveform can often be split up into a chip sequence and a chip waveform

$$g_{k,\mu}(t) = \sum_{\nu=-\infty}^{+\infty} \psi_k(t) * s_{k,\mu}[\nu]\, \delta(t - \nu T_\mathrm{c}) \,.$$

ν: chip time

T_c: chip clock cycle

$\psi_k(t)$: chip waveform

$s_{k,\mu}[\nu]$: chip sequence

In practice, often

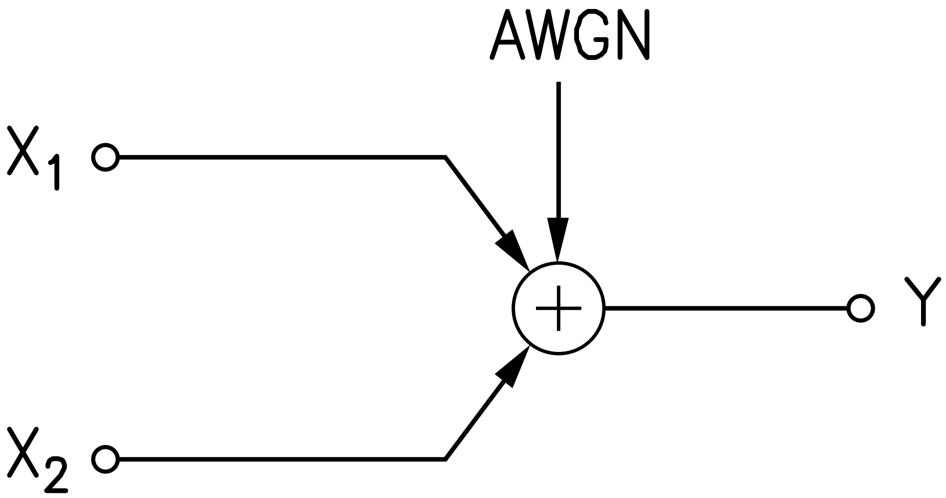
$$\psi_k(t) = \psi(t) \quad \forall\, k$$

$$s_{k,\mu}[\nu] = s_k[\nu] \quad \forall\, \mu$$

Note that chip-asynchronism can be modelled by different chip waveforms amongst the users.

Gaussian Multiple–Access Channel

Two users:

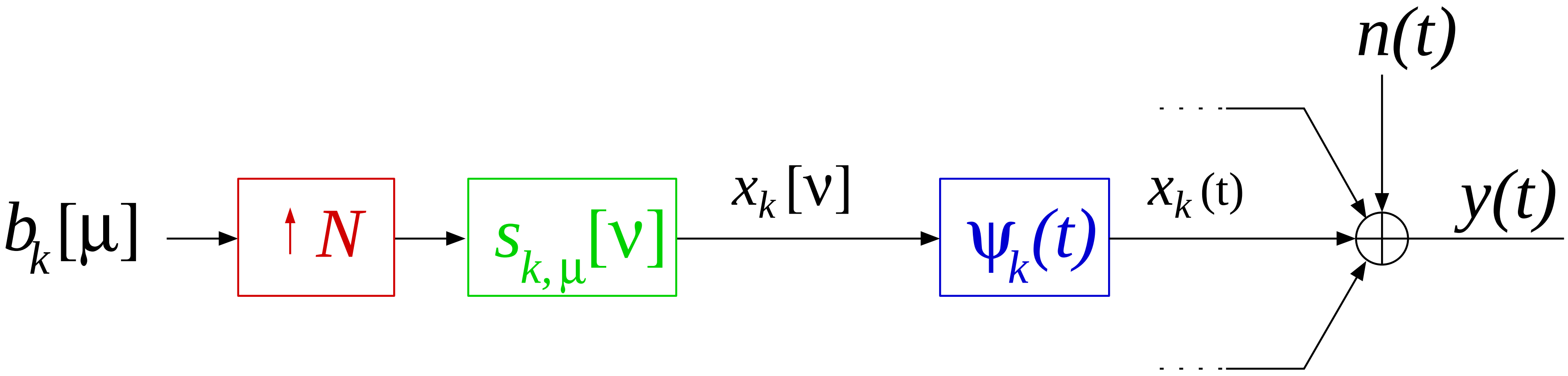


Channel is additive.

Noise is additive, white, and Gaussian distributed.

Noise is independent of X_1 and X_2.

Block Structure of Linear CDMA

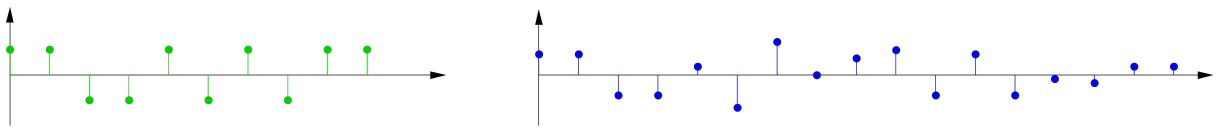


- Upsampling
- Filtering
- Pulse shaping

CDMA with ISI

Construct a set of virtual spreading sequences which is the convolution of the actual spreading sequences and the impulse responses of the channels.

$$s_k[\nu] * h_k[\nu] = \tilde{s}_k[\nu]$$



CDMA with ISI and the actual sequences is equivalent to CDMA with**out** ISI and the virtual sequences.

Though, the actual sequences can be designed orthogonal, the virtual sequences can**not**, unless all channels are known to all users.

For purpose of analysis, interchip interference is often neglected and lost orthogonality is taken into account by the random spreading assumption.

Discrete–Time Channel

If the chip waveforms are identical for all users, i.e. $\psi_k(t) = \psi(t)$, and T_s is a multiple of T_c, there exists a sufficient discrete–time description

$$y[\nu] = n[\nu] + \sum_{k=1}^{K} x_k[\nu]$$

with

$$x_k[\nu] = \sum_{\mu=-\infty}^{+\infty} s_{k,\mu}[\nu - N\mu] b_k[\mu]$$

where

$$N = \frac{T_\mathrm{s}}{T_\mathrm{c}}$$

is called the spreading factor (spreading gain, processing gain).

The discrete–time noise process $n[\nu]$ is white, if $\psi(t)$ is a $\sqrt{\text{Nyquist}}$ waveform.

Discrete–Time Vector Channel

Write sequences as vectors:

$$\underbrace{\begin{bmatrix} y[0] \\ y[1] \\ \vdots \\ y[N-1] \\ y[N+0] \\ y[N+1] \\ \vdots \end{bmatrix}}_{\boldsymbol{y}} = \underbrace{\begin{bmatrix} n[0] \\ n[1] \\ \vdots \\ n[N-1] \\ n[N+0] \\ n[N+1] \\ \vdots \end{bmatrix}}_{\boldsymbol{n}} + \underbrace{\begin{bmatrix} s_{1,0}[0] & \dots & s_{K,0}[0] & s_{1,1}[0-N] & \dots & s_{K,1}[0-N] & \dots \\ s_{1,0}[1] & \dots & s_{K,0}[1] & s_{1,1}[1-N] & \dots & s_{K,1}[1-N] & \dots \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \\ s_{1,0}[N-1] & \dots & s_{K,0}[N-1] & s_{1,1}[-1] & \dots & s_{K,1}[-1] & \dots \\ s_{1,0}[N+0] & \dots & s_{K,0}[N+0] & s_{1,1}[0] & \dots & s_{K,1}[0] & \dots \\ s_{1,0}[N+1] & \dots & s_{K,0}[N+1] & s_{1,1}[1] & \dots & s_{K,1}[1] & \dots \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots \end{bmatrix}}_{\boldsymbol{S}} \underbrace{\begin{bmatrix} b_1[0] \\ \vdots \\ b_K[0] \\ b_1[1] \\ \vdots \\ b_K[1] \\ \vdots \end{bmatrix}}_{\boldsymbol{b}}$$

Equation accounts for asynchronous (but chip–synchronous) users as well as sequences with more than N non–zero chips.

Memoryless and Synchronous Case

Assume

$$s_{k,\mu}[\nu] = 0 \qquad \forall\, \nu < 0, \nu \geq N, k, \mu.$$

Then, the spreading matrix $\boldsymbol{S}$ becomes block–diagonal:

$$\begin{bmatrix} \boldsymbol{y}[0] \\ \boldsymbol{y}[1] \\ \vdots \end{bmatrix} = \begin{bmatrix} \boldsymbol{n}[0] \\ \boldsymbol{n}[1] \\ \vdots \end{bmatrix} + \begin{bmatrix} \boldsymbol{S}[0] & \boldsymbol{0} & \\ \boldsymbol{0} & \boldsymbol{S}[1] & \boldsymbol{0} \\ & \boldsymbol{0} & \ddots \end{bmatrix} \begin{bmatrix} \boldsymbol{b}[0] \\ \boldsymbol{b}[1] \\ \vdots \end{bmatrix}$$

$$\underset{N \times 1}{\boldsymbol{y}[\mu]} = \underset{N \times 1}{\boldsymbol{n}[\mu]} + \underset{N \times K}{\boldsymbol{S}[\mu]} \; \underset{K \times 1}{\boldsymbol{b}[\mu]}$$

Notation:

Discrete time runs in chips and symbols for scalars and vectors (matrices), respectively.

Orthogonal Multiple–Access

Necessary and sufficient condition for orthogonal signals in discrete time:

$$\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S} \quad \text{is diagonal}$$

Time–division multiple–access:

$$\boldsymbol{S} \quad \text{is (weighted) permutation matrix}$$

Orthogonal frequency–division multiplexing:

$$\boldsymbol{S}[\mu] \quad \text{is (part of) FFT matrix}$$

Orthogonal code–division multiple–access:

$$\boldsymbol{S}[\mu] \quad \text{is (part of) Walsh–Hadamard matrix}$$

Correlated Waveforms

Why correlated waveforms?

- More users than spreading factor, i.e. $K > N$.
- No synchronism required.
- Some channels destroy orthogonality anyway.

Popular design of spreading sequences:

- Pseudo–noise sequences, e.g. maximum-length sequences
- Gold sequences
- Kasami sequences

Random Waveforms

Each spreading sequence is chosen randomly.

Popular model for purpose of performance analysis.

In the large–system limit, i.e. $K, N \to \infty$, analytical expressions are known for the singular values of the spreading matrix $\boldsymbol{S}$.

Exist for both $K > N$ and $K \leq N$.

The marginal probability distribution of the chips hardly matters. It is irrelevant in the large system limit in many cases.

Multi–User Detection

The problem:

- Given the observation $y(t)$, find the most likely transmitted sequence of data vectors $\boldsymbol{b}[\mu]$.
- Special case of a vector–classification problem
- In general, **np–complete**, i.e. it belongs to a class of problems for which no algorithm is known whose complexity scales as a polynomial function of K.
- In particular cases, multi–user detection is not np–complete, e.g. orthogonal sequences or maximum-length sequences.

Sufficient Discrete–Time Statistics

Theorem 1 *The outputs of a bank of K linear filters matched to the K symbol waveforms form a set of sufficient statistics for estimation of all users' data in AWGN. If the chip waveform is unique to all users and $\sqrt{\text{Nyquist}}$, they can be sampled at the symbol rate.*

$$\begin{aligned} \boldsymbol{v} &= \boldsymbol{A}^{-1}\boldsymbol{S}^{\mathrm{H}}\boldsymbol{y} \\ &= \boldsymbol{A}^{-1}\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S}\boldsymbol{b} + \boldsymbol{A}^{-1}\boldsymbol{S}^{\mathrm{H}}\boldsymbol{n} \end{aligned}$$

The matrix $\boldsymbol{A}$ is arbitrary, but invertible.

Note that a unique chip waveform implicitly assumes chip-synchronous reception.

Memoryless and Synchronous Case

Unless stated otherwise, all further considerations assume this case.

$$\boldsymbol{v}[\mu] = \boldsymbol{A}^{-\mathrm{H}}[\mu]\boldsymbol{S}^{\mathrm{H}}[\mu]\boldsymbol{S}[\mu]\boldsymbol{b}[\mu] + \boldsymbol{A}^{-\mathrm{H}}[\mu]\boldsymbol{S}^{\mathrm{H}}[\mu]\boldsymbol{n}[\mu]$$

Define the cross–correlation matrix

$$\boldsymbol{R}[\mu] \stackrel{\triangle}{=} \boldsymbol{A}^{-\mathrm{H}}[\mu]\boldsymbol{S}^{\mathrm{H}}[\mu]\boldsymbol{S}[\mu]\boldsymbol{A}^{-1}[\mu].$$

Then, the CDMA channel is canonically described by

$$\boldsymbol{v}[\mu] = \boldsymbol{R}[\mu]\boldsymbol{A}[\mu]\boldsymbol{b}[\mu] + \boldsymbol{A}^{-\mathrm{H}}[\mu]\boldsymbol{S}^{\mathrm{H}}[\mu]\boldsymbol{n}[\mu].$$

Further Assumptions

Unless stated otherwise, all further considerations are conditioned on the following assumptions:

- All random processes are ergodic.
- The data symbols are statistically independent in both μ and k.
- The users have power
$$\boldsymbol{A}^2 \triangleq \operatorname{diag}\left(\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S}\right).$$
- The noise power is σ^2.

The dependency on μ is not always stated explicitly.

Optimum Receiver

Definition 3 *The jointly optimum receiver minimizes*

$$\Pr\left(\hat{\boldsymbol{b}} \neq \boldsymbol{b} \middle| \boldsymbol{v}\right).$$

Definition 4 *The individually optimum receiver minimizes*

$$\Pr\left(\hat{b}_k \neq b_k \middle| \boldsymbol{v}\right) \qquad \forall k.$$

Lemma 1 *The individually optimum receiver minimizes*

$$\sum_{k=1}^{K} \Pr\left(\hat{b}_k \neq b_k \middle| \boldsymbol{v}\right).$$

Output Distribution

The probability density function (PDF) of the sufficient statistics given $\boldsymbol{b}$

$$\mathrm{p}_{\boldsymbol{v}|\boldsymbol{b}}(\boldsymbol{v},\boldsymbol{b}) = \frac{\exp\left(-\frac{1}{\sigma^2}\left(\boldsymbol{v}-\boldsymbol{RAb}\right)^{\mathrm{H}}\boldsymbol{R}^{-1}\left(\boldsymbol{v}-\boldsymbol{RAb}\right)\right)}{(\pi\sigma^2)^K \det \boldsymbol{R}}$$

is actually the distribution of the matched–filtered noise.

Joint PDF of sufficient statistics and data is

$$\mathrm{p}_{\boldsymbol{v},\boldsymbol{b}}(\boldsymbol{v},\boldsymbol{b}) = \mathrm{p}_{\boldsymbol{v}|\boldsymbol{b}}(\boldsymbol{v},\boldsymbol{b})\mathrm{p}_{\boldsymbol{b}}(\boldsymbol{b})$$

with

$$\mathrm{p}_{\boldsymbol{b}}(\boldsymbol{b}) = \sum_{\tilde{\boldsymbol{b}}\in\mathcal{A}^K} \Pr\left(\tilde{\boldsymbol{b}}\right)\delta\left(\boldsymbol{b}-\tilde{\boldsymbol{b}}\right).$$

Jointly Optimum A–Posteriori Detection

Bayesian law gives:

$$\mathrm{p}_{\boldsymbol{b}|\boldsymbol{v}}(\boldsymbol{v},\boldsymbol{b}) = \frac{\mathrm{p}_{\boldsymbol{v}|\boldsymbol{b}}(\boldsymbol{v},\boldsymbol{b})\mathrm{p}_{\boldsymbol{b}}(\boldsymbol{b})}{\int \mathrm{p}_{\boldsymbol{v}|\boldsymbol{b}}\left(\boldsymbol{v},\tilde{\boldsymbol{b}}\right)\mathrm{p}_{\boldsymbol{b}}\left(\tilde{\boldsymbol{b}}\right)\mathrm{d}\tilde{\boldsymbol{b}}}$$

Denominator is irrelevant for detection:

$$\begin{aligned}
\operatorname*{argmax}_{\boldsymbol{b}} \mathrm{p}_{\boldsymbol{b}|\boldsymbol{v}}(\boldsymbol{v},\boldsymbol{b}) &= \operatorname*{argmax}_{\boldsymbol{b}} \mathrm{p}_{\boldsymbol{v}|\boldsymbol{b}}(\boldsymbol{v},\boldsymbol{b})\mathrm{p}_{\boldsymbol{b}}(\boldsymbol{b}) \\
&= \operatorname*{argmax}_{\boldsymbol{b}\in\mathcal{A}^K} \mathrm{p}_{\boldsymbol{v}|\boldsymbol{b}}(\boldsymbol{v},\boldsymbol{b})\Pr(\boldsymbol{b}) \\
&= \operatorname*{argmax}_{\boldsymbol{b}\in\mathcal{A}^K} -\frac{1}{\sigma^2}(\boldsymbol{v}-\boldsymbol{R}\boldsymbol{A}\boldsymbol{b})^{\mathrm{H}}\boldsymbol{R}^{-1}(\boldsymbol{v}-\boldsymbol{R}\boldsymbol{A}\boldsymbol{b}) + \log\Pr(\boldsymbol{b}) \\
&= \operatorname*{argmin}_{\boldsymbol{b}\in\mathcal{A}^K} \boldsymbol{b}^{\mathrm{H}}\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S}\boldsymbol{b} - 2\Re\boldsymbol{v}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{b} - \sigma^2\log\Pr(\boldsymbol{b})
\end{aligned}$$

Individually Optimum A–Posteriori Detection

Bayesian law gives:

$$\mathrm{p}_{b_k|\boldsymbol{v}}(\boldsymbol{v}, b_k) = \frac{\int \mathrm{p}_{\boldsymbol{v}|\boldsymbol{b}}\left(\boldsymbol{v}, \tilde{\boldsymbol{b}}\right) \mathrm{p}_{\boldsymbol{b}}\left(\tilde{\boldsymbol{b}}\right) \prod_{i \neq k} \mathrm{d}\tilde{b}_i}{\int \mathrm{p}_{\boldsymbol{v}|\boldsymbol{b}}\left(\boldsymbol{v}, \tilde{\boldsymbol{b}}\right) \mathrm{p}_{\boldsymbol{b}}\left(\tilde{\boldsymbol{b}}\right) \mathrm{d}\tilde{\boldsymbol{b}}}$$

Denominator is irrelevant for detection:

$$\begin{aligned} \underset{b_k}{\operatorname{argmax}}\, \mathrm{p}_{b_k|\boldsymbol{v}}(\boldsymbol{v}, b_k) &= \underset{b_k}{\operatorname{argmax}} \int \mathrm{p}_{\boldsymbol{v}|\boldsymbol{b}}\left(\boldsymbol{v}, \tilde{\boldsymbol{b}}\right) \mathrm{p}_{\boldsymbol{b}}\left(\tilde{\boldsymbol{b}}\right) \prod_{i \neq k} \mathrm{d}\tilde{b}_i \\ &= \underset{b_k \in \mathcal{A}}{\operatorname{argmax}} \sum_{\tilde{\boldsymbol{b}} \in \mathcal{A}^K : \tilde{b}_k = b_k} \mathrm{p}_{\boldsymbol{v}|\boldsymbol{b}}\left(\boldsymbol{v}, \tilde{\boldsymbol{b}}\right) \Pr\left(\tilde{\boldsymbol{b}}\right) \end{aligned}$$

Maximum Likelihood Detection

If all bits are transmitted equally likely, i.e.

$$\Pr(\boldsymbol{b}) = |\mathcal{A}|^{-K}$$

the detection rules slightly simplifies:

- jointly optimum detection

$$\operatorname*{argmin}_{\boldsymbol{b}\in\mathcal{A}^K} \boldsymbol{b}^{\mathrm{H}}\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S}\boldsymbol{b} - 2\Re\boldsymbol{v}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{b}$$

- individually optimum detection

$$\operatorname*{argmax}_{b_k\in\mathcal{A}} \sum_{\tilde{\boldsymbol{b}}\in\mathcal{A}^K:\tilde{b}_k=b_k} \mathrm{p}_{\boldsymbol{v}|\boldsymbol{b}}\left(\boldsymbol{v},\tilde{\boldsymbol{b}}\right)$$

Linear Multi–User Detection

Since the optimum detectors are np–complete, suboptimum approaches are frequently used in practice.

Definition 5 *A multiuser detector is called linear if its estimate is formed by component–wise quantization of a linear transform on the sufficient statistics.*

$$\hat{\boldsymbol{b}} = \underset{\mathcal{A}}{\mathrm{quant}} \left(\boldsymbol{L}\boldsymbol{v} \right)$$

with

$$\underset{\mathcal{A}}{\mathrm{quant}}(x) \stackrel{\triangle}{=} \underset{\tilde{x}\in\mathcal{A}}{\mathrm{argmin}} \, |x - \tilde{x}|$$

Single–User Matched Filter

The SUMF (conventional detector) ignores the presence of multi–user interference:

$$\boldsymbol{L} = \boldsymbol{A}^{-1}$$

- Used for sake of simplicity
- Poor performance

Decorrelator

The decorrelator follows from the approximation

$$\mathcal{A}^K \approx \mathbb{C}^K$$

and equal prior probability for all symbols

$$\mathrm{p}_{\boldsymbol{b}}(\boldsymbol{b}) = \lim_{\xi \to \infty} \prod_{k=1}^{K} \begin{cases} (\pi\xi)^{-1} & \text{for} \quad |b_k| < \xi \\ 0 & \text{otherwise} \end{cases}.$$

It is given as

$$\boldsymbol{L} = \boldsymbol{A}^{-1} \boldsymbol{R}^{-1}.$$

For signal sets with constant amplitude, it does not depend on the users' powers.

LMMSE Detector

The LMMSE detector follows from the approximation

$$\mathcal{A}^K \approx \mathbb{C}^K$$

and a Gaussian density for the transmitted symbols

$$\mathrm{p}_{\boldsymbol{b}}(\boldsymbol{b}) = \pi^{-K} \exp\left(-\boldsymbol{b}^{\mathrm{H}} \boldsymbol{b}\right)$$

It minimizes the mean squared error

$$\mathrm{E}\left(\boldsymbol{b} - \hat{\boldsymbol{b}}\right)^{\mathrm{H}} \left(\boldsymbol{b} - \hat{\boldsymbol{b}}\right)$$

among all linear detectors and is given by the filter matrix

$$\boldsymbol{L} = \boldsymbol{A}^{-1}\left(\boldsymbol{R} + \sigma^2 \boldsymbol{A}^{-2}\right)^{-1}.$$

For vanishing noise, it becomes identical to the decorrelator.

For overwhelming noise, it becomes identical to the SUMF.

Unbiased LMMSE Detector

The bias problem:
For orthogonal sequences, the LMMSE detector is worse than the SUMF.

This is overcome by constraining the LMMSE detector to

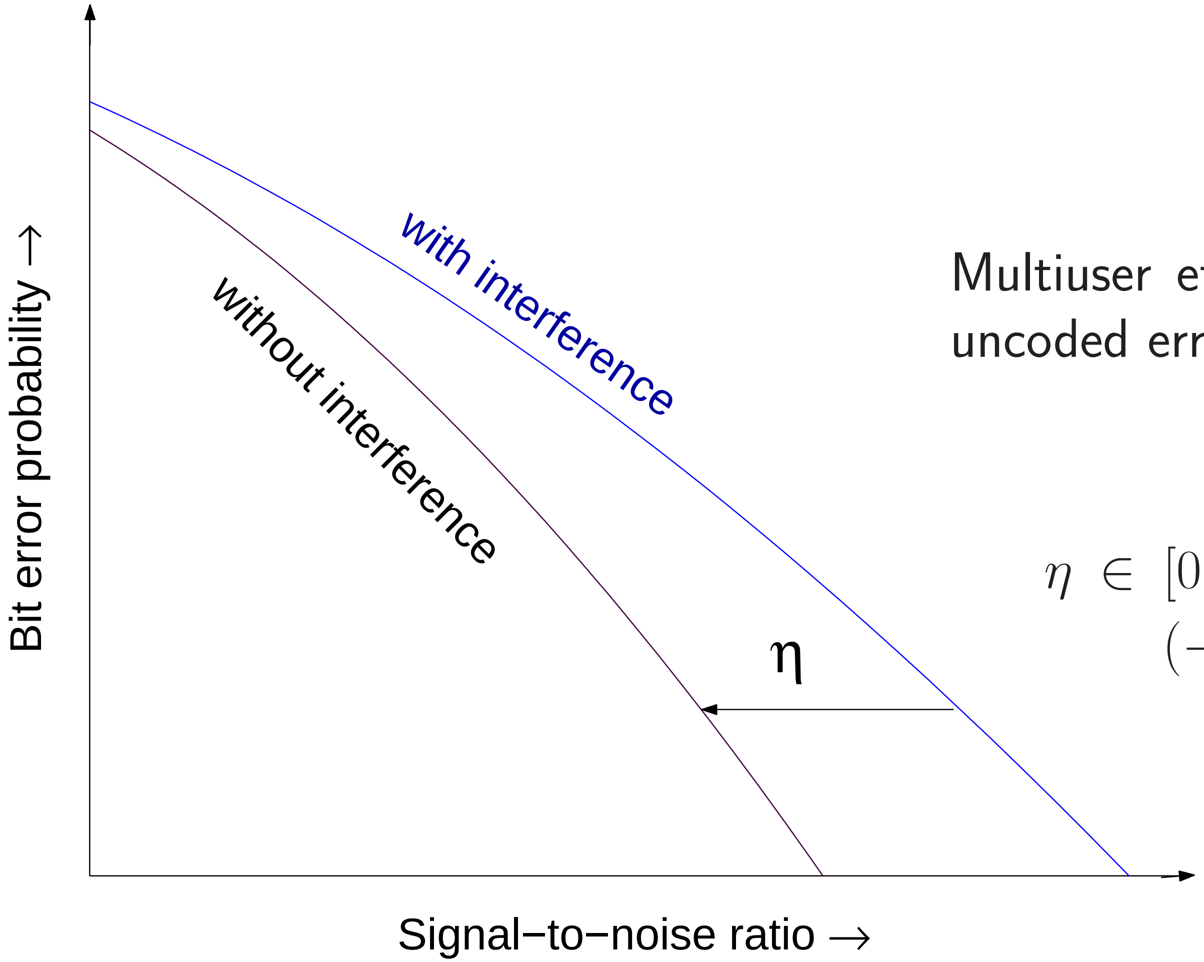
$$\operatorname{diag}(\boldsymbol{LRA}) = \mathbf{I}.$$

This leads to the detector

$$\boldsymbol{L} = \operatorname{diag}^{-1}\left(\boldsymbol{A}^2 + \sigma^2 \boldsymbol{R}^{-1}\right)^{-1} \boldsymbol{A}^{-3} \left(\boldsymbol{R} + \sigma^2 \boldsymbol{A}^{-2}\right)^{-1}$$

For signal sets with constant amplitude, the bias problem does not occur.

Multi-User Efficiency



Multiuser efficiency measures uncoded error probability.

$$\eta \in [0;1]$$
$$(-\infty \text{ dB}; 0 \text{ dB}]$$

Multi-User Efficiency *(cont'd)*

Definition 6 *Let $P_k(\sigma^2, \boldsymbol{R}, \text{'det'})$ denote the uncoded symbol error rate of user k after detection with detector 'det' and signature sequences with covariance matrix $\boldsymbol{R}$. Then, the number η_k ensuring*

$$P_k(\sigma^2, \boldsymbol{R}, \text{'det'}) = P_k(\sigma^2/\eta_k, \mathbf{I}, \text{'MAP'})$$

is called multi–user efficiency of user k with detector 'det'.

The multi–user efficiency lies within

$$\eta_k \in [0; 1].$$

The asymptotic multi–user efficiency is given as

$$\tilde{\eta}_k = \lim_{\sigma \to 0} \eta_k.$$

Verdú introduced the notion of multiuser efficiency in 1986.

Sergio Verdú
born in Barcelona in 1958

The Decorrelator

Assume $\boldsymbol{R}$ is invertible:
The decorrelator completely suppresses all interference, but enhances the AWGN.

$$\eta_k = \frac{1}{\left(\boldsymbol{R}^{-1}\right)_{kk}}$$

Since, multi–user efficiency is independent of AWGN and interfering users' powers:

$$\tilde{\eta}_k = \eta_k$$

Unless, the cross–correlation matrix is singular, multi–user efficiency is non–zero.

The LMMSE Detector

The LMMSE detector maximizes the signal-to-interference-and-noise ratio (SINR) among all linear detectors.

$$\mathrm{SINR}_k = \boldsymbol{s}_k^{\mathrm{H}} \left(\boldsymbol{S}\boldsymbol{S}^{\mathrm{H}} - \boldsymbol{s}_k \boldsymbol{s}_k^{\mathrm{H}} + \sigma^2 \mathbf{I}\right)^{-1} \boldsymbol{s}_k = \frac{\boldsymbol{s}_k^{\mathrm{H}} \left(\boldsymbol{S}\boldsymbol{S}^{\mathrm{H}} + \sigma^2 \mathbf{I}\right)^{-1} \boldsymbol{s}_k}{1 - \boldsymbol{s}_k^{\mathrm{H}} \left(\boldsymbol{S}\boldsymbol{S}^{\mathrm{H}} + \sigma^2 \mathbf{I}\right)^{-1} \boldsymbol{s}_k}$$

Multi–user efficiency is approximately

$$\eta_k \approx \mathrm{SINR}_k \, \frac{\sigma^2}{A_k^2}.$$

This is only an approximation as the remaining interference plus noise is, in general, not exactly Gaussian distributed.

The LMMSE Detector *(cont'd)*

Consider the eigenvalue decomposition

$$\boldsymbol{S}\boldsymbol{S}^{\mathrm{H}} - \boldsymbol{s}_k\boldsymbol{s}_k^{\mathrm{H}} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{H}}.$$

Then,

$$\mathrm{SINR}_k = \tilde{\boldsymbol{s}}_k^{\mathrm{H}} \left(\boldsymbol{\Lambda} + \sigma^2\mathbf{I}\right)^{-1} \tilde{\boldsymbol{s}}_k$$

with

$$\tilde{\boldsymbol{s}}_k = \boldsymbol{V}^{\mathrm{H}}\boldsymbol{s}_k$$

The performance of the LMMSE detector depends on the eigenvalue distribution of $\boldsymbol{S}\boldsymbol{S}^{\mathrm{H}} - \boldsymbol{s}_k\boldsymbol{s}_k^{\mathrm{H}}$ and the transformed spreading sequence $\tilde{\boldsymbol{s}}_k$.

The aim of this course is to study the properties of the two for several wireless communication channels.

The Stieltjes Transform

The densities for most other projections cannot be given in explicit form. They are more easily characterized in terms of their Stieltjes transforms

$$G(s) \triangleq \int \frac{f(x)dx}{x-s} \qquad \Im(s) > 0.$$

Stieltjes Inversion Formula:

$$f(x) = \lim_{y \to 0+} \frac{1}{\pi} \Im\left[G(x+\mathrm{j}y)\right]$$

Convergence of the LMMSE-SINR

Consider the linear MMSE detector studied in Chapter 1 with a real-valued spreading matrix $\boldsymbol{S}$.

$$\mathrm{SINR}_k = \tilde{\boldsymbol{s}}_k^{\mathrm{T}} \left(\boldsymbol{\Lambda} + \sigma^2 \mathbf{I}\right)^{-1} \tilde{\boldsymbol{s}}_k.$$

We have the almost sure convergence

$$\tilde{\boldsymbol{s}}_k^{\mathrm{T}} \left(\boldsymbol{\Lambda} + \sigma^2 \mathbf{I}\right)^{-1} \tilde{\boldsymbol{s}}_k \longrightarrow \mathrm{Tr}\left(\boldsymbol{\Lambda} + \sigma^2 \mathbf{I}\right)^{-1} \qquad \text{with } \mathrm{Tr}(\cdot) = \lim_{K\to\infty} \frac{1}{K} \mathrm{trace}(\cdot).$$

From the deformed quarter circle law, we get the almost sure identity

$$\mathrm{Tr}\left(\sigma^2 \mathbf{I} + \boldsymbol{\Lambda}\right)^{-1} = \mathrm{G}_{\boldsymbol{\Lambda}}\left(-\sigma^2\right).$$

Thus,

$$\mathrm{SINR}_k \longrightarrow \frac{(1-\alpha)P}{2\sigma^2} - \frac{1}{2} + \sqrt{\frac{(1-\alpha)^2 P^2}{4\sigma^4} + \frac{(1+\alpha)P}{2\sigma^2} + \frac{1}{4}}.$$

The SINR is almost surely identical for all users.

LMMSE Detector with Random Spreading

Theorem 7 *Let the chips of any user be i.i.d. zero–mean random variables with finite sixth moment and the sequences of all users jointly independent. Then, the multi–user efficiencies of all users converge almost surely, as $N, K \to \infty$ but*

$$\alpha \triangleq \frac{K}{N}$$

fixed, to the deterministic unique positive solution of the fixed–point equation

$$\frac{1}{\eta_{\mathrm{LMMSE}}} = 1 + \alpha \int \frac{x}{\sigma^2 + \eta_{\mathrm{LMMSE}}\, x} \,\mathrm{dP}_{A^2}(x),$$

if the powers of the users converge weakly to the limit distribution $\mathrm{P}_{A^2}(x)$.

The multi–user efficiency in large systems is **identical** for all users regardless of their powers.

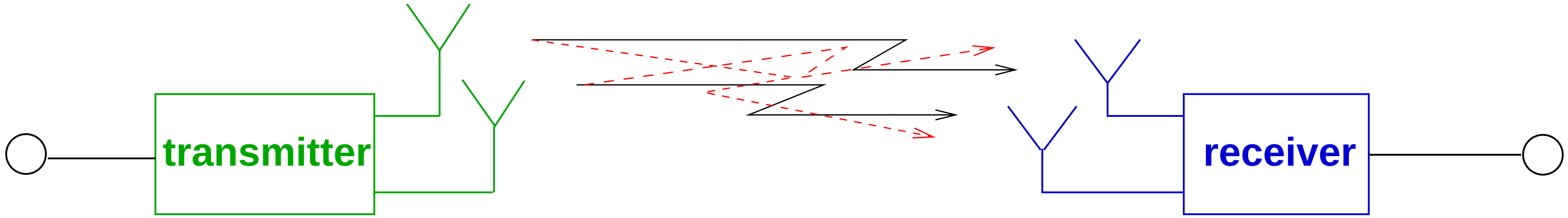
Chapter 3:

Antenna Arrays

Dual Antenna Arrays

Consider a single user communication system with T antenna elements at transmitter site and R antenna elements at receiver site.

Example ($T = R = 2$):



Channel is described by

$$\boldsymbol{y}[\mu] = \boldsymbol{n}[\mu] + \boldsymbol{H}[\mu]\boldsymbol{b}[\mu]$$

with $\boldsymbol{H}$ containing the TR channel coefficients from the T transmit to the R receive antennas at discrete time μ.

Dual Antenna Arrays as Special Case of CDMA

Regard

the antenna elements at transmitter site as users indexed by k

and

the antenna elements at receiver site as discrete chips at "time" (space) instant ν.

Then, dual antenna arrays become equivalent to CDMA with spreading matrix

$$\boldsymbol{S}[\mu] = \boldsymbol{H}[\mu].$$

A vector of sufficient statistics can be formed by (spatial) matched filtering

$$\boldsymbol{v}[\mu] = \boldsymbol{A}^{-1}[\mu]\boldsymbol{H}^{\mathrm{H}}[\mu]\Big(\boldsymbol{H}[\mu]\boldsymbol{b}[\mu] + \boldsymbol{n}[\mu]\Big)$$

The standard algorithms of multi–user detection apply without changes.

I.i.d. Complex Gaussian Fading

Assume the entries of $\boldsymbol{H}$ are i.i.d. complex Gaussian.

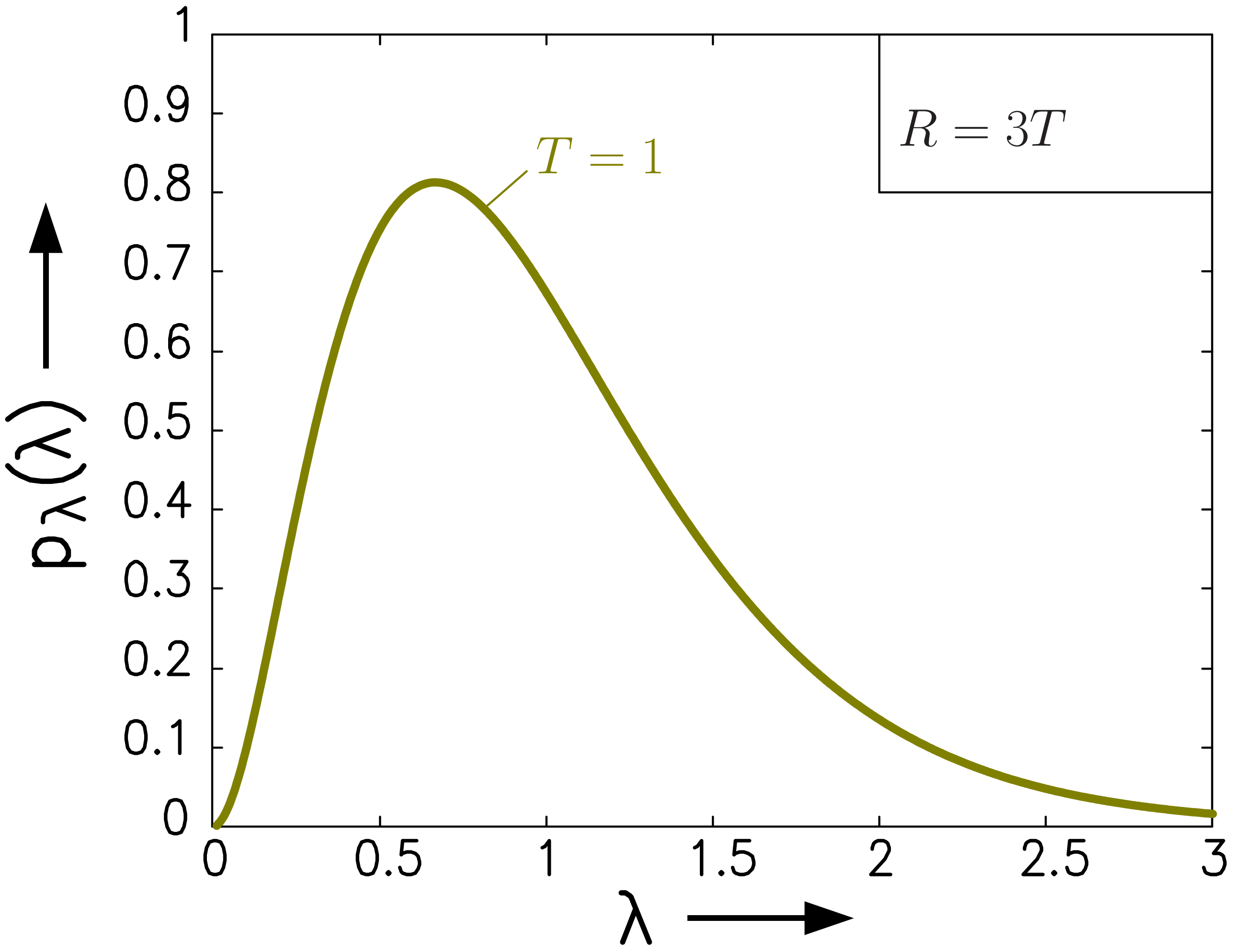
Then, the eigenvalues of $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}$ are distributed as

$$\mathrm{p}_\lambda(x) = \begin{cases} \dfrac{R}{T} \displaystyle\sum_{k=0}^{\min\{T,R\}-1} \frac{k!}{(k+|T-R|)!} \left(\mathrm{L}_k^{(|T-R|)}(xR)\right)^2 (xR)^{|T-R|} \mathrm{e}^{-xR} & \text{for } x > 0 \\ \left(1-\frac{R}{T}\right) \begin{cases} 0 & \text{for } T \le R \\ \delta(x) & \text{for } T > R \end{cases} & \text{otherwise} \end{cases}$$

with the Laguerre polynomials

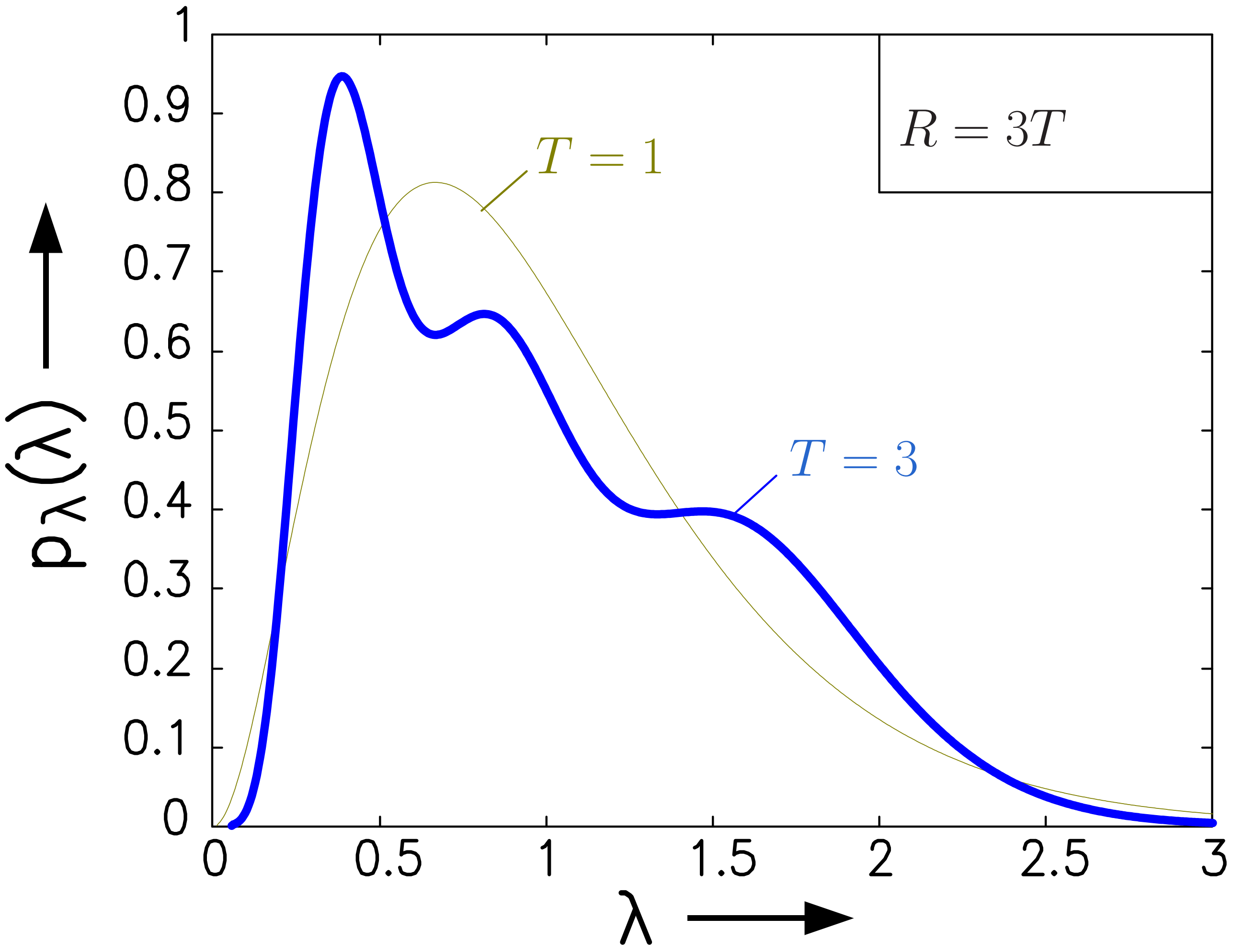
$$\mathrm{L}_a^{(b)}(x) \triangleq \frac{\mathrm{e}^x}{a!\,x^b} \frac{\mathrm{d}^a}{\mathrm{d}x^a}\left(x^{a+b}\,\mathrm{e}^{-x}\right)$$

I.i.d. Complex Gaussian Fading (cont'd)



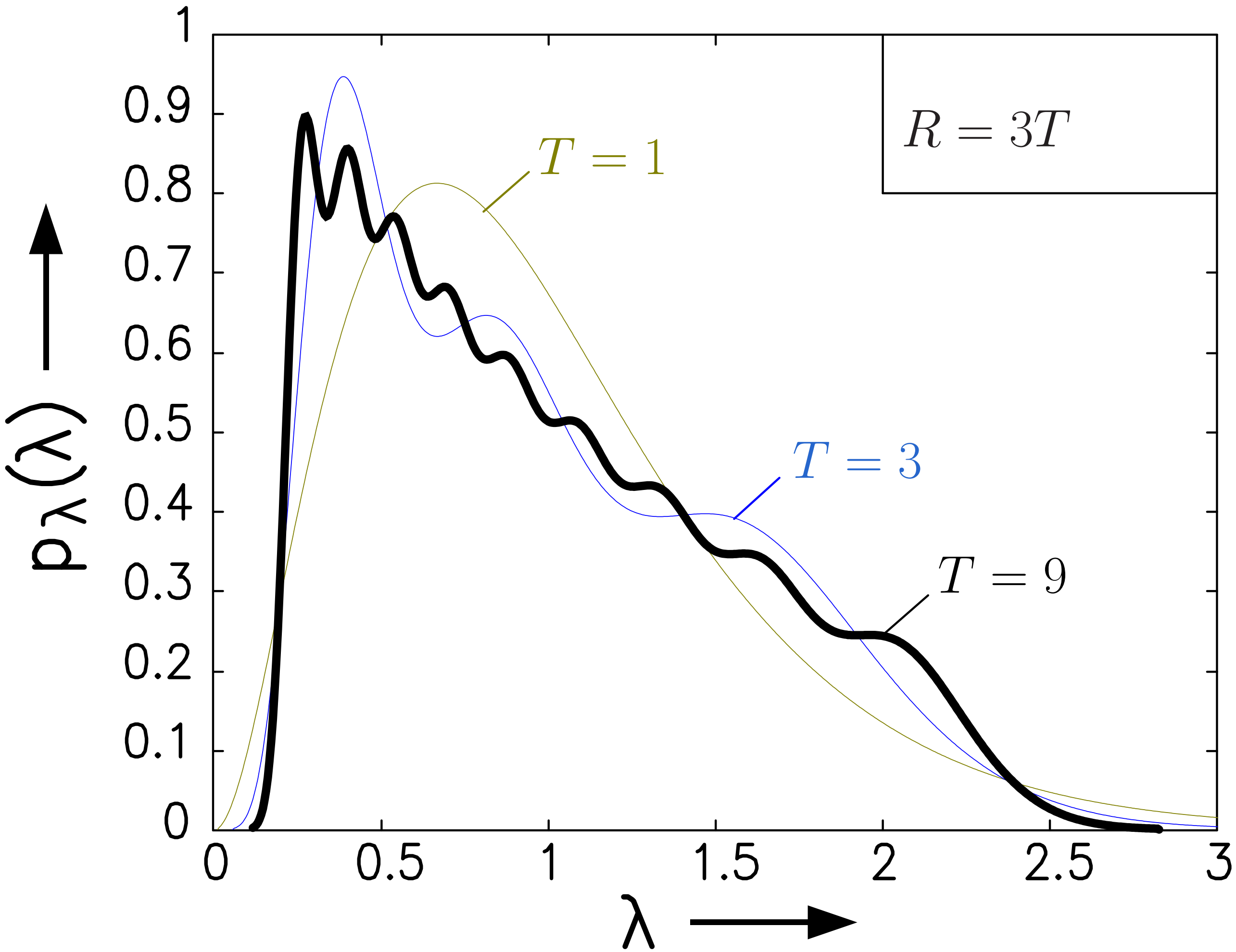
Eigenvalues of $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}$ for i.i.d. entries in the $R \times T$ matrix $\boldsymbol{H}$.

I.i.d. Complex Gaussian Fading (cont'd)



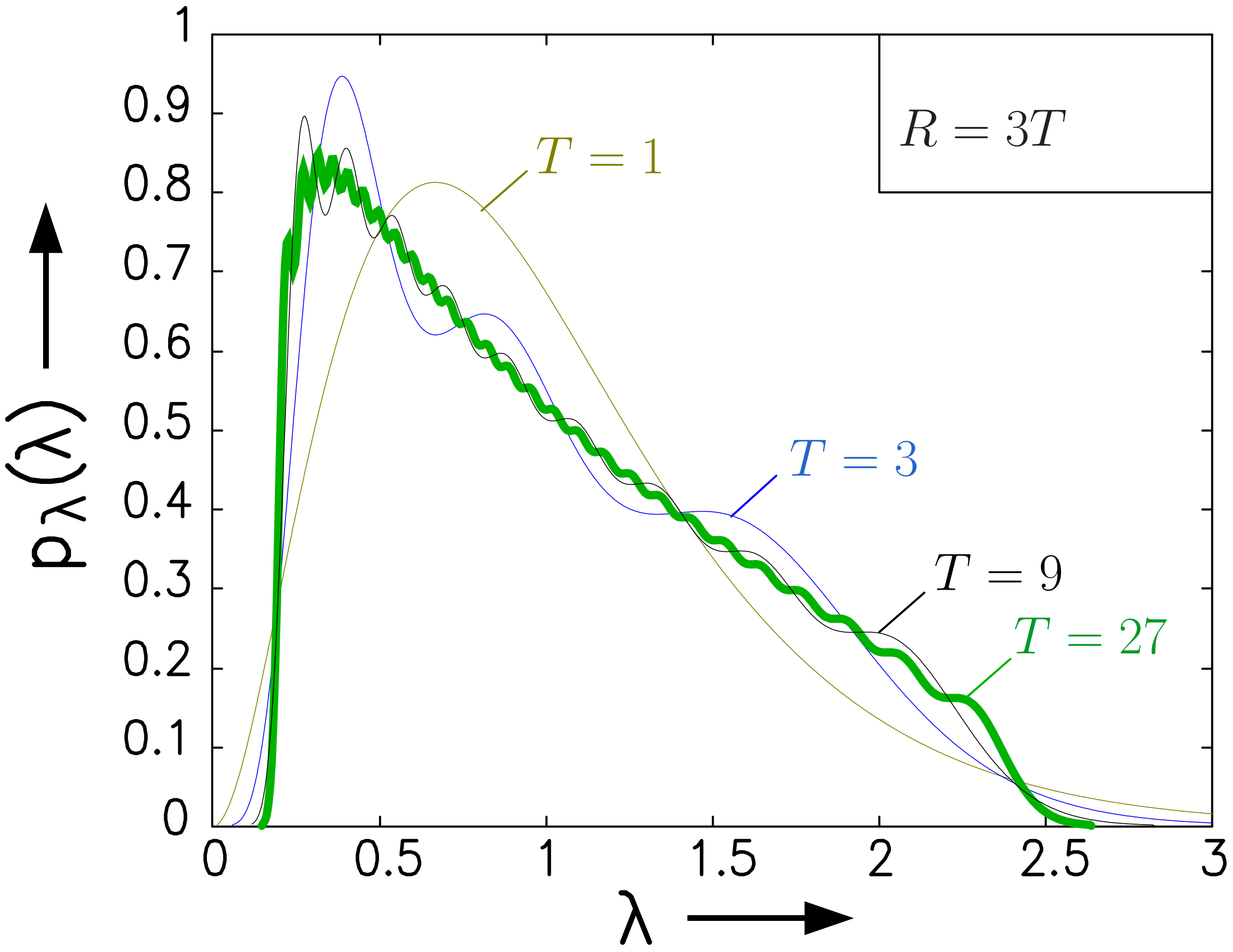
Eigenvalues of $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}$ for i.i.d. entries in the $R \times T$ matrix $\boldsymbol{H}$.

I.i.d. Complex Gaussian Fading (cont'd)



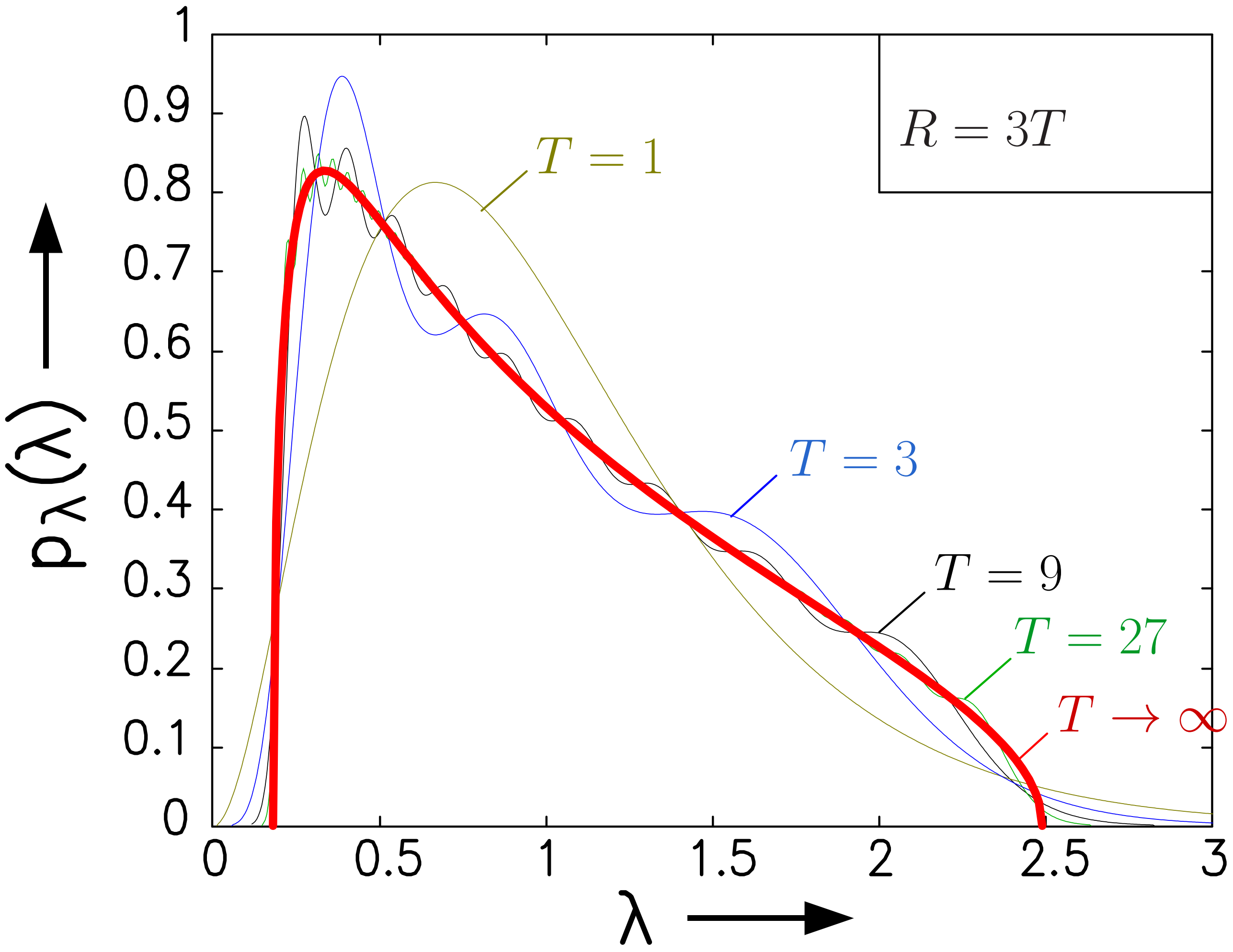
Eigenvalues of $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}$ for i.i.d. entries in the $R \times T$ matrix $\boldsymbol{H}$.

I.i.d. Complex Gaussian Fading (cont'd)



Eigenvalues of $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}$ for i.i.d. entries in the $R \times T$ matrix $\boldsymbol{H}$.

I.i.d. Complex Gaussian Fading (cont'd)



Eigenvalues of $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}$ for i.i.d. entries in the $R \times T$ matrix $\boldsymbol{H}$.

CDMA with Dual Antenna Arrays

Without loss of generality $T = K$.

The system is described by the virtual $NR \times K$ spreading matrix

$$\tilde{\boldsymbol{S}} = \begin{bmatrix} h_{11}\boldsymbol{s}_1 & h_{12}\boldsymbol{s}_2 & \dots & h_{1K}\boldsymbol{s}_K \\ h_{21}\boldsymbol{s}_1 & h_{22}\boldsymbol{s}_2 & \dots & h_{2K}\boldsymbol{s}_K \\ \vdots & \vdots & \ddots & \vdots \\ h_{R1}\boldsymbol{s}_1 & h_{R2}\boldsymbol{s}_2 & \dots & h_{RK}\boldsymbol{s}_K \end{bmatrix}$$

Note that with the Kronecker product $\otimes$:

$$\tilde{\boldsymbol{s}}_k = \boldsymbol{h}_k \otimes \boldsymbol{s}_k$$

Note also that the entries of $\tilde{\boldsymbol{S}}$ are not jointly independent even if those ones of $\boldsymbol{S}$ and $\boldsymbol{H}$ are.

A Resource Pooling Result

Theorem 8 *Let the chips of any user be i.i.d. zero–mean complex Gaussian random variables, the sequences of all users jointly independent, and the antenna array channel* h_{rk} *follow the i.i.d. complex Gaussian model. Then, the multi–user efficiency of the linear MMSE detector converges for all users almost surely, as* $N, K \to \infty$ *but* $\alpha \triangleq \frac{K}{N}$ *and* R *fixed, to the deterministic unique positive solution of the fixed–point equation*

$$\frac{1}{\eta_{\mathrm{LMMSE}}} = 1 + \frac{\alpha}{R} \int \frac{x}{\sigma^2 + \eta_{\mathrm{LMMSE}}\, x} \mathrm{dP}_{\tilde{A}^2}(x),$$

if the powers of the users converge weakly to the limit distribution $\mathrm{P}_{\tilde{A}^2}(x)$ *with*

$$|\tilde{A}_k|^2 = |A_k|^2 \sum_{r=1}^{R} |h_{rk}|^2.$$

Factor i.i.d. Model

Trouble of the i.i.d. model:
Dependencies among entries of $\boldsymbol{H}$ due to

- limited number of scatterers
- correlation between closely spaced antennas

Factor channel model

$$\underset{R \times T}{\boldsymbol{H}} = \underset{R \times S}{\boldsymbol{\Phi}} \; \underset{S \times T}{\boldsymbol{\Theta}}$$

can be confirmed by measurements for i.i.d. matrices $\boldsymbol{\Phi}$ and $\boldsymbol{\Theta}$, and appropriate choice of S.

For $S \to \infty$, the entries of $\boldsymbol{H}$ become i.i.d.

For the factor channel model, large system results for several detectors are known.

Kronecker Model

Trouble of the i.i.d. model:
Dependencies among entries of $\boldsymbol{H}$ due to

- limited number of scatterers
- correlation between closely spaced antennas

Kronecker channel model

$$\underset{R \times T}{\boldsymbol{H}} = \underset{R \times R}{\sqrt{\boldsymbol{C}_R}} \underset{R \times T}{\boldsymbol{G}} \underset{T \times T}{\sqrt{\boldsymbol{C}_T}}$$

can be confirmed by measurements for an i.i.d. complex Gaussian matrix $\boldsymbol{G}$ and appropriate choices for the correlation matrices $\boldsymbol{C}_R$ and $\boldsymbol{C}_T$.

For the Kronecker channel model, large system results for several detectors are known.

Jointly Correlated Channel Model

Trouble of the i.i.d. model:
Dependencies among entries of $\boldsymbol{H}$ due to

- limited number of scatterers
- correlation between closely spaced antennas

Jointly correlated channel model

$$\underset{RT \times RT}{\boldsymbol{C}} = \mathrm{E}\ \underset{RT \times 1}{\mathrm{vec}(\boldsymbol{H})^{\mathrm{H}}}\ \underset{1 \times RT}{\mathrm{vec}(\boldsymbol{H})}$$

The Kronecker model is a special case with $\boldsymbol{C} = \boldsymbol{C}_T \otimes \boldsymbol{C}_R$.

Some measurements agree only with the jointly correlated channel model.

Generalized Kronecker Channel Model

Kronecker channel model:

$$\begin{array}{ccccccc} \boldsymbol{H} & = & \sqrt{\boldsymbol{C}_R} & \boldsymbol{\Phi} & \boldsymbol{A} & \boldsymbol{\Theta} & \sqrt{\boldsymbol{C}_T} \\ R \times T & & R \times R & R \times S & S \times S & S \times T & T \times T \end{array}$$

The matrices $\boldsymbol{\Phi}$ and $\boldsymbol{\Theta}$ are steering matrices. They depend on the array geometry and the location of scattering objects.

The matrices $\boldsymbol{C}_R$ and $\boldsymbol{C}_T$ are coupling matrices. They depend only on the array geometry. They converge to identity matrices for large element spacing.

The steering matrices can be well approximated by i.i.d. random matrices. With this assumption the product $\boldsymbol{\Phi A \Theta}$ converges to an i.i.d. random matrix for $S \to \infty$ (rich scattering).

Why look steering matrices like i.i.d. random matrices?

Pseudo-Randomness in Steering Matrices

Example: Uniform linear array, scatterers in far-field

$$\mathbf{\Theta}_{s,t} = \exp(\mathrm{j}\vartheta_{s,t}) = \exp\left(\mathrm{j}\theta_s - \mathrm{j}(t-1)\frac{2\pi d}{\lambda}\sin(\alpha_s)\right)$$

Linear congruential random number generator (used in MATLAB up to version 4)

$$X_{n+1} = (aX_n + c) \bmod m, \qquad n \geq 0$$

with seed X_0 and $a = 1$.

Each scattering object acts as random number generator with its distance as seed and the sine of its angle times the element spacing as increment.

Correlated Resource Pooling

Theorem 9 *Let the chips of any user be i.i.d. zero–mean complex Gaussian random variables, the sequences of all users jointly independent, and the empirical distributions of the channel gains h_{rk} across the users converge, jointly for all receive antennas r to an R-dimensional joint limit distribution $\mathrm{P}_{\boldsymbol{H}}(x)$. Then, with linear MMSE detection, the SINR of user k converges, as $N, K \to \infty$ but $\alpha \triangleq \frac{K}{N}$ and R fixed, conditioned on the channel gains of user k to*

$$\frac{\boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{A} \boldsymbol{h}_k}{\sigma^2}$$

where $\boldsymbol{A}$ is the deterministic unique positive definite solution of the matrix-valued fixed–point equation

$$\boldsymbol{A}^{-1} = \mathbf{I} + \alpha \int \frac{\boldsymbol{x}\boldsymbol{x}^{\mathrm{H}}}{\sigma^2 + \boldsymbol{x}^{\mathrm{H}} \boldsymbol{A} \boldsymbol{x}} \, \mathrm{dP}_{\boldsymbol{H}}(\boldsymbol{x}),$$

Asymptotic performance is characterized by an $R \times R$ matrix.

Channel Capacity

Assume that the channel is known to the receiver, but unknown to the transmitter. Then, the channel capacity per transmit antenna is given as

$$C = \frac{1}{T} \log\det\left(\mathbf{I} + \frac{1}{\sigma^2}\boldsymbol{H}^\dagger\boldsymbol{H}\right).$$

Note that for $T \to \infty$

$$\begin{aligned}\frac{\partial C}{\partial \sigma^2} &= \frac{1}{\sigma^2}\mathrm{Tr}\left(\mathbf{I} + \frac{1}{\sigma^2}\boldsymbol{H}^\dagger\boldsymbol{H}\right)^{-1} - \frac{1}{\sigma^2} \\ &= \frac{1}{\sigma^2}\mathrm{G}_{\boldsymbol{H}^\dagger\boldsymbol{H}}(-\sigma^2) - \frac{1}{\sigma^2}.\end{aligned}$$

The channel capacity can be obtained more easily from the Stieltjes transform than from the eigenvalue density.

Chapter 4:

Low-Complexity Multiuser Detection

Single–User Matched Filter

Theorem 10 *Let the chips of any user be i.i.d. zero–mean random variables with finite fourth moment and the sequences of all users jointly independent. Then, the multi–user efficiencies of all users converge almost surely, as $N, K \to \infty$ but $\alpha = \frac{K}{N}$ fixed, to*

$$\eta_{\mathrm{SUMF}} = \frac{1}{1 + \alpha \int \frac{x}{\sigma^2} \, \mathrm{dP}_{A^2}(x)},$$

if the powers of the users converge weakly to the limit distribution $\mathrm{P}_{A^2}(x)$.

Large system approximation:

$$\mathsf{SINR}_k = \frac{A_k^2}{\sigma^2 + \frac{1}{N} \sum_{\substack{i=1 \\ i \neq k}}^{K} A_i^2}$$

Linear Parallel Interference Cancellation (LPIC)

This is a linear receiver in terms of Definition 5.

Goal:
Reduce effort for signal–processing at expense of performance.

Two stages for $\boldsymbol{A} = \mathbf{I}$:

$$\hat{\boldsymbol{b}} = \underset{\mathcal{A}}{\text{quant}} \left(\boldsymbol{v} - (\boldsymbol{R} - \mathbf{I})\boldsymbol{v} \right)$$

Three stages:

$$\hat{\boldsymbol{b}} = \underset{\mathcal{A}}{\text{quant}} \Big(\boldsymbol{v} - (\boldsymbol{R} - \mathbf{I})\big(\boldsymbol{v} - (\boldsymbol{R} - \mathbf{I})\boldsymbol{v}\big) \Big)$$

D stages:

$$\hat{\boldsymbol{b}} = \underset{\mathcal{A}}{\text{quant}} \left(\sum_{i=0}^{D-1} (\mathbf{I} - \boldsymbol{R})^i \boldsymbol{v} \right)$$

LPIC (cont'd)

$$\boldsymbol{L}_{\mathrm{LPIC},D} = \sum_{i=0}^{D-1} (\mathbf{I} - \boldsymbol{R})^i .$$

If $\lambda_{\max}(\boldsymbol{R}) < 2$ and $\lambda_{\min}(\boldsymbol{R}) > 0$, then

$$\boldsymbol{L}_{\mathrm{LPIC},\infty} = \boldsymbol{R}^{-1}.$$

Theorem 11 *Let the chips of any user be i.i.d. zero–mean random variables with finite variance and the sequences of all users jointly independent. Then, the largest and smallest eigenvalue of $\boldsymbol{R}$ converge almost surely to*

$$\left(1+\sqrt{\alpha}\right)^2 \qquad \text{and} \qquad \left(1-\sqrt{\alpha}\right)^2,$$

respectively, as $N, K \to \infty$ but $\alpha = \frac{K}{N}$ fixed.

Convergence holds for random spreading if

$$\alpha < \left(\sqrt{2}-1\right)^2 \approx 0.17$$

Weighted LPIC

Let $\boldsymbol{R}$ be non–singular.

Let λ_i denote the eigenvalues of $\boldsymbol{R}$.

Then,

$$\prod_{k=1}^{K} \left(\boldsymbol{R} - \lambda_k \mathbf{I} \right) = \mathbf{0} \qquad \Longrightarrow \qquad -\mathbf{I} + \sum_{k=1}^{K} \alpha_k \boldsymbol{R}^k = \mathbf{0}$$

Cayley–Hamilton Theorem — with appropriate α_k s.

Solution to matrix inversion problem given the eigenvalues:

$$\boldsymbol{R}^{-1} = \sum_{k=1}^{K} \alpha_k \boldsymbol{R}^{k-1}$$

Weighted LPIC (cont'd)

Linear MMSE filter: $$\boldsymbol{L}_{\text{MMSE}} = \left(\boldsymbol{R} + \sigma^2 \mathbf{I}\right)^{-1}$$

Approximation by power series:

Cayley–Hamilton theorem yields:

$$\left(\boldsymbol{R} + \sigma^2 \mathbf{I}\right)^{-1} = \sum_{i=0}^{K-1} \tilde{w}_i \boldsymbol{R}^i$$
$$\approx \sum_{i=0}^{D-1} w_i \boldsymbol{R}^i \qquad \text{for } D < K.$$

For random spreading the optimum weights converge almost surely, as $K, N \to \infty$ with $\alpha = \frac{K}{N}$, and can be given in closed form.

Semi-Universal Weights

Filter shall be independent from the realization of the random matrix $\boldsymbol{S}$, but may use its statistics.

For most large random matrices, as $K = \alpha N \to \infty$, many finite dimensional functions of the eigenvalues, e.g. the filter coefficients, freeze.

The asymptotic limits depend only on parts of the statistics of the random matrix.

The weights can be calculated off-line with the help of random matrix and free probability theory.

Weight Design

is given by Yule–Walker equations:

$$\begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_{D+1} \end{bmatrix} = \begin{bmatrix} m_2 + \sigma^2 m_1 & m_3 + \sigma^2 m_2 & \dots & m_{D+2} + \sigma^2 m_{D+1} \\ m_3 + \sigma^2 m_2 & m_4 + \sigma^2 m_3 & \dots & m_{D+3} + \sigma^2 m_{D+2} \\ \vdots & \vdots & \ddots & \vdots \\ m_{D+2} + \sigma^2 m_{D+1} & m_{D+3} + \sigma^2 m_{D+2} & \dots & m_{2D+2} + \sigma^2 m_{2D+1} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_D \end{bmatrix}$$

with the total moments

$$m_n \stackrel{\triangle}{=} \mathrm{E}\left\{\lambda^n\right\} = \mathrm{Tr}\left(\boldsymbol{S}^{\mathrm{H}}\boldsymbol{S}\right)^n$$

Example for Asymptotic Weight Design

Random matrix with i.i.d. entries:

$$m_n = \frac{1}{n}\sum_{i=1}^{n}\binom{n}{i}\binom{n}{i-1}\alpha^i.$$

$D=2$	$w_0 = -\sigma^2 w_1 + 2 + 2\alpha$ $w_1 = -1$
$D=3$	$w_0 = -\sigma^2 w_1 + 3 + 4\alpha + 3\alpha^2$ $w_1 = -\sigma^2 w_2 - 3 - 3\alpha$ $w_2 = 1$
$D=4$	$w_0 = -\sigma^2 w_1 + 4 + 6\alpha + 6\alpha^2 + 4\alpha^3$ $w_1 = -\sigma^2 w_2 - 6 - 9\alpha - 6\alpha^2$ $w_2 = -\sigma^2 w_3 + 4 + 4\alpha$ $w_3 = -1$
$D=5$	$w_0 = -\sigma^2 w_1 + 5 + 8\alpha + 9\alpha^2 + 8\alpha^3 + 5\alpha^4$ $w_1 = -\sigma^2 w_2 - 10 - 18\alpha - 18\alpha^2 - 10\alpha^3$ $w_2 = -\sigma^2 w_3 + 10 + 16\alpha + 10\alpha^2$ $w_3 = -\sigma^2 w_4 - 5 - 5\alpha$ $w_4 = 1$

Rate of Convergence

Theorem 12 *Let $\boldsymbol{A} = \mathbf{I}$ and the chips of any user be i.i.d. zero–mean random variables with finite fourth moment and the sequences of all users jointly independent. Then, the multi–user efficiencies of all users converge almost surely, as $N, K \to \infty$ but $\alpha = \frac{K}{N}$ fixed, to*

$$\eta_{\mathrm{WLPIC},D+1} = \frac{1}{1 + \dfrac{\alpha}{\sigma^2 + \eta_{\mathrm{WLPIC},D}}}$$

with $\eta_0 = 0$ for optimally chosen weights.

The approximation converges to the exact MMSE performance as a *continued fraction*.

For optimal coefficients w_i, the approximation error ϵ decreases exponentially with the number of stages D:

$$\epsilon < \mathrm{const.}\,(1 + \mathrm{SNR})^{-D}$$

There are even tighter bounds.

Individual Weight Design

Allow for different weights for different users

$$\left(\boldsymbol{R} + \sigma^2 \mathbf{I}\right)^{-1} = \sum_{i=0}^{K-1} \tilde{w}_i \boldsymbol{R}^i$$

$$\approx \sum_{i=0}^{D-1} \boldsymbol{W}_i \boldsymbol{R}^i \qquad \text{for } D < K \qquad \text{and all } \boldsymbol{W}_i \text{ diagonal.}$$

Weight design by the same Yule-Walker equations, but with the k-partial moments

$$m_n^{(k)} = \left[\left(\boldsymbol{S}^{\mathrm{H}} \boldsymbol{S} \right)^n \right]_{kk} .$$

For users with different powers, individual weight design is better.

Do the k-partial moments convergence asymptotically?

Convergence of Partial Moments

Let the random matrix $\boldsymbol{S}$ fulfill the same conditions as needed for the deformed quarter circle law. Let $\boldsymbol{A}$ be a $K \times K$ diagonal matrix such that its singular value distribution converges almost surely, as $K \to \infty$ to a non-random limit distribution. Let

$$\boldsymbol{R} = \boldsymbol{A}^{\mathrm{H}} \boldsymbol{S}^{\mathrm{H}} \boldsymbol{S} \boldsymbol{A}.$$

Then, $\left(\boldsymbol{R}^{\ell}\right)_{kk}$, the k^{th} diagonal element of $\boldsymbol{R}^{\ell}$ converges, conditioned on a_{kk}, the k^{th} diagonal element of $\boldsymbol{A}$, almost surely, as $K = \alpha N \to \infty$ to

$$R_{kk}^{(\ell)} = |a_{kk}|^2 \alpha \sum_{q=1}^{\ell} R_{kk}^{(q-1)} m_{\ell-q}^{(\boldsymbol{R})}, \qquad \ell > 1$$

with

$$m_q^{(\boldsymbol{R})} = \mathrm{Tr}(\boldsymbol{R}^q) = \lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} R_{kk}^{(q)}.$$

Correlated Resource Pooling

Let

- $\boldsymbol{S} = [\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_K]$ with $\boldsymbol{s}_k = \boldsymbol{h}_k \otimes \tilde{\boldsymbol{s}}_k$ where $\tilde{\boldsymbol{S}}$ is i.i.d.
- the entries of $\boldsymbol{H}$ may be arbitrarily dependent as long as the rows have a joint limit distribution and are finite in number.

Then, as the dimensions of $\tilde{\boldsymbol{S}}$ grow

- the k-partial moments conditioned on $\boldsymbol{h}_k$ converge and
- recursive expressions for them are known.

Multipath Fading Channels

Let the path differences be only a few chip intervals. Approximate the linear time shift by a cyclic shift modulo N. For large N this becomes more and more accurate.

2 paths: All odd column of the $N \times 2K$ matrix $\boldsymbol{S}$ are i.i.d. Each even column of $\boldsymbol{S}$ is a cyclically shifted version of the adjacent column to the left.

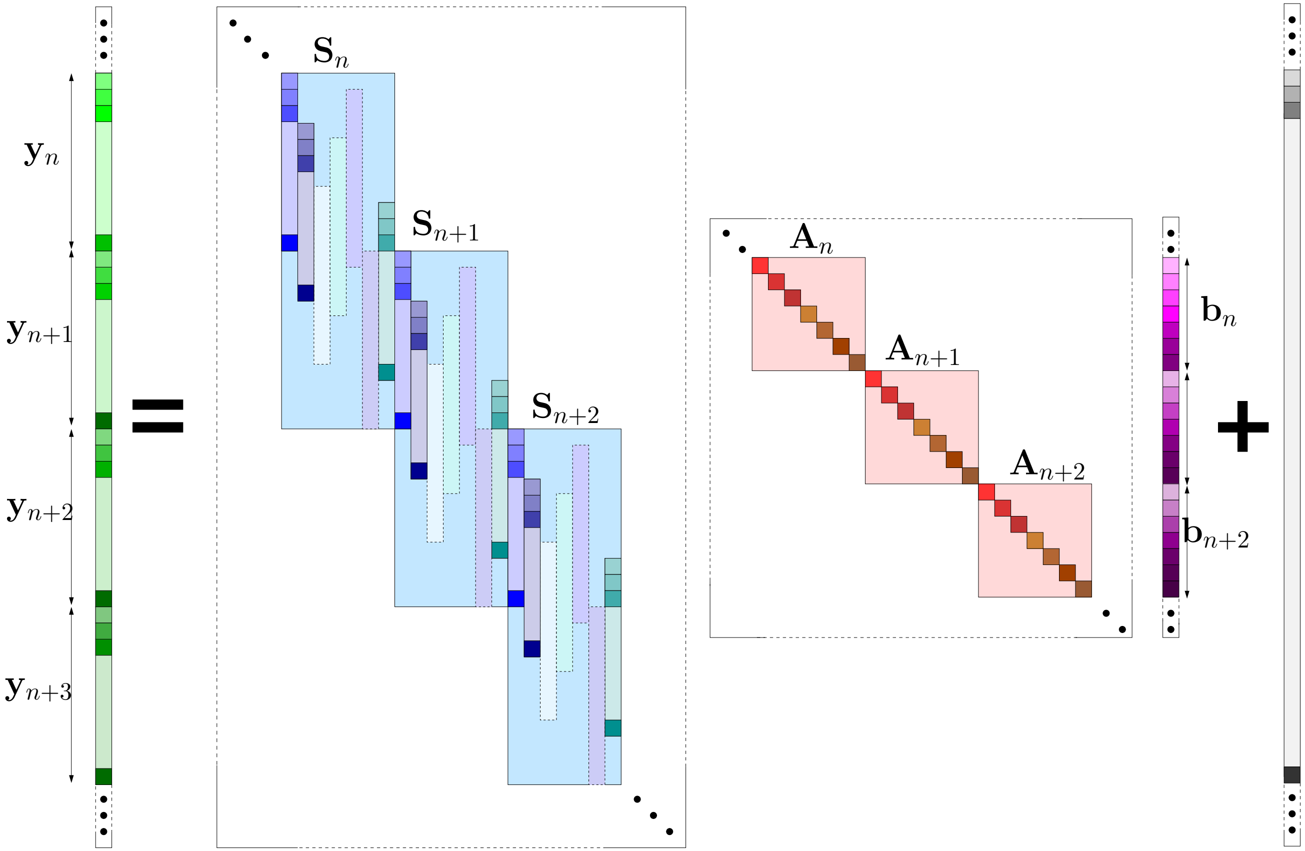
$$\mathrm{E}\left\{\boldsymbol{b}\boldsymbol{b}^{\mathrm{H}}\right\} = \mathbf{I} \otimes \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

and $(\boldsymbol{A})_{kk}$ are independent zero-mean and complex Gaussian.

This setting is equivalent to the full i.i.d. setting in all asymptotic aspects if the users' powers follow the same distribution.

Equivalence holds for an arbitrary number of paths.

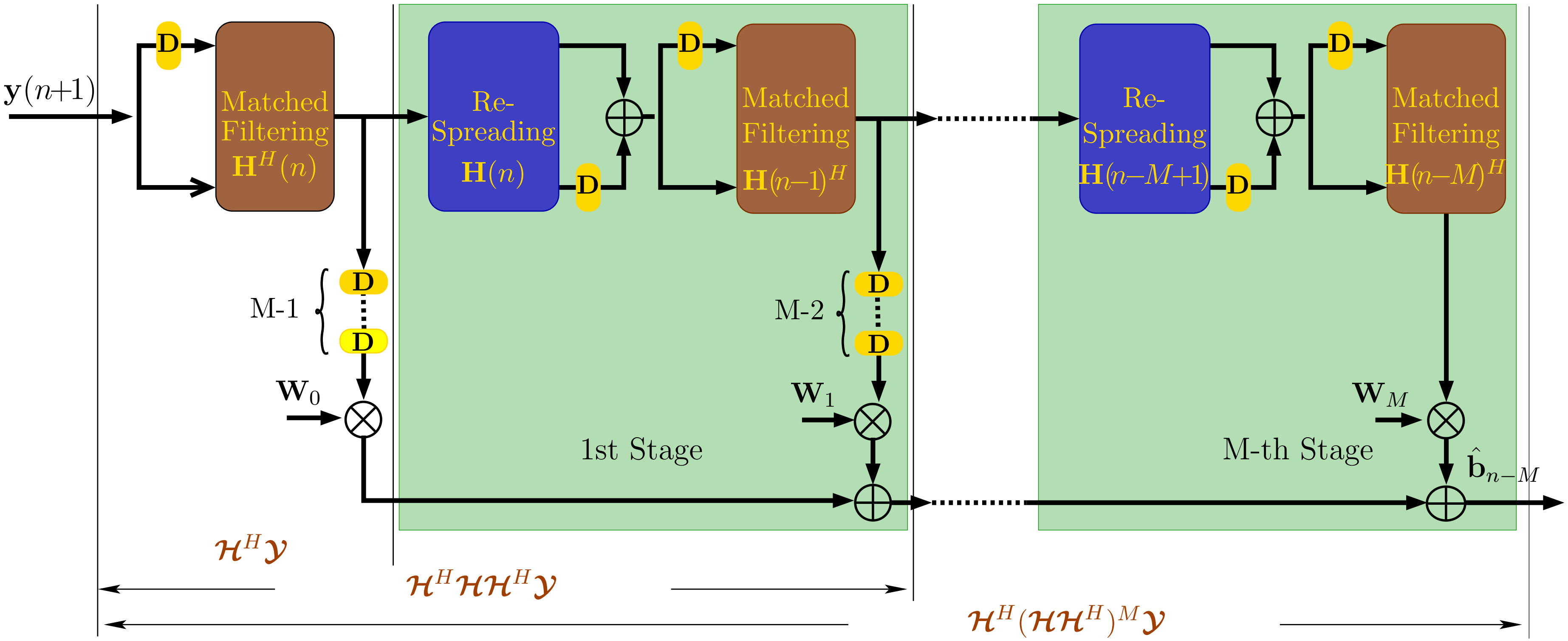
Asynchronous Users



Convergence of k-partial moments proven, recursive expressions to construct them known.

$$\mathcal{Y} = \underbrace{\mathcal{S}\mathcal{A}}_{\mathcal{H}}\mathcal{B} + \mathcal{N}$$

Detector Structure for Asynchronous Users



No truncation effects.

Chip-Asynchronous Users

Let the delays of the users be spread out over a chip interval.

The excess bandwidth of the chip waveforms can be utilized to span signal dimensions.

Improvements in

- SINR of single user matched filter
- SINR of linear multiuser receivers
- Total capacity per chip

Fixed point equations to characterize the large-system performance are known.

Recursive expressions for partial moments are known (depend on delay distribution).

De-synchronization on the chip level improves performance.

Chip-Asynchronous LMMSE Detection

Theorem 13 ([10]) *Let the spreading sequences of any user be i.i.d. zero mean Gaussian random variables, the users be chip-asynchronous with uniformly distributed relative delay, the users' powers be independent of the delays and converge to the limit distribution* $\mathrm{P}_{A^2}(x)$*. Let the chip waveform have spectrum* $\Psi(\omega)$ *and unit energy and the noise be white with spectral density* N_0*. Then, the multiuser efficiency of the linear MMSE detector converges in probability as* $K, N \to \infty$ *with* $\frac{K}{N} \to \alpha$ *to*

$$\eta = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \eta\left(\omega\right) \mathrm{d}\omega$$

where the multiuser efficiency spectral density $\eta\left(\omega\right)$ *is the unique solution to the fixed point equation*

$$\frac{1}{\eta\left(\omega\right)} = \frac{T_\mathrm{c}}{\left|\Psi\left(\omega\right)\right|^2} + \alpha \int \frac{x \mathrm{d}\mathrm{P}_{A^2}(x)}{N_0 + \eta x T_\mathrm{c}}.$$

This requires over-sampling to form sufficient discrete-time statistics.

Additive Free Convolution

Let $\boldsymbol{A} = \boldsymbol{A}^{\mathrm{H}}$ and $\boldsymbol{B} = \boldsymbol{B}^{\mathrm{H}}$ be free and

$$\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B}.$$

Then,

$$\mathrm{R}_{\boldsymbol{C}}(w) = \mathrm{R}_{\boldsymbol{A}}(w) + \mathrm{R}_{\boldsymbol{B}}(w)\,.$$

The R-transform is defined as

$$\mathrm{R}(w) \triangleq \mathrm{G}^{-1}(-w) - \frac{1}{w}$$

where $\mathrm{G}(\cdot)$ denotes the Stieltjes transform and $\mathrm{G}^{-1}(\cdot)$ its inverse with respect to composition.

The R-transform linearizes additive free convolution.

Effective Interference

Large system approximation for the SINR of the LMMSE detector for a finite number of users:

$$\mathsf{SINR}_k \approx \frac{A_k^2}{\sigma_n^2 + \frac{1}{N}\sum\limits_{\substack{i=1\\i\neq k}}^{K} \frac{A_i^2}{1+\mathsf{SINR}_i}}$$

Effective interference from user i:

$$I_i = \frac{1}{N} \cdot \frac{A_i^2}{1+\mathsf{SINR}_i}$$

Definition of R-transform:

$$\mathrm{G}(s) = \frac{1}{-s + \mathrm{R}(-\mathrm{G}(s))}$$

R-transform for equal power users:

$$\mathrm{R}_{\mathsf{EP}}(w) = \frac{\alpha}{1-w}$$

R-transform for power profile:

$$\mathrm{R}_{\mathsf{PP}}(w) = \int \frac{\alpha P \mathrm{dP}(P)}{1-wP}$$

Effective interferences are the R-transforms of all individual users.

Is There a Free Log-Normal Distribution?

Consider the channel matrix

$$\boldsymbol{H}_N = \boldsymbol{M}_N \boldsymbol{M}_{N-1} \cdots \boldsymbol{M}_2 \boldsymbol{M}_1 \triangleq \prod_{n=1}^{N} \boldsymbol{M}_n$$

where the matrices $\boldsymbol{M}_1$, $\boldsymbol{M}_{1<n<N}$, and $\boldsymbol{M}_N$ denote the subchannels from the transmitter array to the first cluster of scatterers, from the $(n-1)^{\mathrm{st}}$ cluster of scatterers to the n^{th} cluster, and from the $(N-1)^{\mathrm{st}}$ cluster to the receiving array, respectively. Let all matrices $\boldsymbol{M}_n$ have size $K \times K$.

We ask for the asymptotic eigenvalue distribution of the matrix $\boldsymbol{C}_N \triangleq \boldsymbol{H}_N \boldsymbol{H}_N^{\mathrm{H}}$.

Assume that the family $\left(\{\boldsymbol{M}_1^{\mathrm{H}} \boldsymbol{M}_1\}, \{\boldsymbol{M}_2^{\mathrm{H}} \boldsymbol{M}_2\}, \ldots, \{\boldsymbol{M}_N^{\mathrm{H}} \boldsymbol{M}_N\}\right)$ is asymptotically free as K tends to infinity.

Consider also the random covariance matrices

$$\tilde{\boldsymbol{C}}_N \triangleq \left(\prod_{n=1}^{N-1} \boldsymbol{M}_n\right) \left(\prod_{n=1}^{N-1} \boldsymbol{M}_n\right)^{\mathrm{H}} \boldsymbol{M}_N^{\mathrm{H}} \boldsymbol{M}_N$$

There is No Free Log-Normal Distribution

The asymptotic eigenvalue distribution of $\boldsymbol{C}_N$ is calculated recursively.

For that purpose, note that the eigenvalues of the matrices $\tilde{\boldsymbol{C}}_N$ and $\boldsymbol{C}_N$ are identical.

Let the entries of $M_{1\leq n\leq N}$ be independent and identically distributed with zero-mean and variance $1/K$. Then,

$$\mathrm{S}_{\boldsymbol{C}_N}(z) = \mathrm{S}_{\tilde{\boldsymbol{C}}_N}(z) = \mathrm{S}_{\boldsymbol{C}_{N-1}}(z)\,\mathrm{S}_{\boldsymbol{M}_N^{\mathrm{H}}\boldsymbol{M}_N}(z) = \frac{\mathrm{S}_{\boldsymbol{C}_{N-1}}(z)}{1+z} = \frac{1}{(1+z)^N}.$$

Back to Stieltjes domain

$$s\left(\Upsilon_{\boldsymbol{C}_N}(s)+1\right)^{N+1} = \Upsilon_{\boldsymbol{C}_N}(s)$$

$$s^{-1}\left(-s\mathrm{G}_{\boldsymbol{C}_N}(s)\right)^{N+1} + s\mathrm{G}_{\boldsymbol{C}_N}(s) = -1.$$

For large N, we get

$$\lim_{N\to\infty} \mathrm{G}_{\boldsymbol{C}_N}(s) = \frac{1}{-s}$$

Almost all eigenvalues converge to zero.

Worst Case Power Distribution

Theorem 17 *Let the chips of any user be i.i.d. zero–mean random variables with finite fourth moment, the sequences of all users jointly independent, and the powers of all users bounded from above and below by positive numbers. Moreover, let $N, K \to \infty$ but $\alpha = \frac{K}{N}$ fixed. Then,*

$$\underset{\boldsymbol{A}:\mathrm{tr}\boldsymbol{A}^2/K=P}{\mathrm{argmin}} \quad \eta = P\mathbf{I}$$

holds for the multi–user efficiency of any linear detector that can be written as a matrix polynomial in $\boldsymbol{R}$.

Equal power interferers are the worst case for given total interference power.

Free Fourier Transform

Let $\boldsymbol{Q}$ be any matrix of bounded norm and bounded rank n and let $\boldsymbol{J}$ be free of $\boldsymbol{Q}$. Then,

$$\lim_{K\to\infty} \frac{1}{K} \log \operatorname*{E}_{\boldsymbol{J}} \mathrm{e}^{K \operatorname{trace}(\boldsymbol{J}\boldsymbol{Q})} = \sum_{a=1}^{n} \int\limits_{0}^{\lambda_a(\boldsymbol{Q})} \mathrm{R}_{\boldsymbol{J}}(w) \mathrm{d}w.$$

This implies

$$\begin{aligned} \mathrm{R}_{\boldsymbol{J}}(\boldsymbol{Q}) &= \frac{\partial}{\partial \boldsymbol{Q}} \lim_{K\to\infty} \frac{1}{K} \log \operatorname*{E}_{\boldsymbol{J}} \mathrm{e}^{K \operatorname{trace}(\boldsymbol{J}\boldsymbol{Q})} \\ &= \lim_{K\to\infty} \frac{\operatorname*{E}_{\boldsymbol{J}} \boldsymbol{J}\, \mathrm{e}^{K \operatorname{trace}(\boldsymbol{J}\boldsymbol{Q})}}{\operatorname*{E}_{\boldsymbol{J}} \mathrm{e}^{K \operatorname{trace}(\boldsymbol{J}\boldsymbol{Q})}} \end{aligned}$$

The bounded rank condition can be relaxed to ranks that grow slowlier than $\sqrt{K}$.

Free Similar Random Matrices

Let the $N \times N$ matrices $\boldsymbol{X}_j, \forall j$, be such that their inverses $\boldsymbol{X}_j^{-1}, \forall j$, exist and

$$\mathrm{Tr}\left(\boldsymbol{X}_i \boldsymbol{X}_j^{-1}\right) = 0 \qquad \forall i \neq j$$

holds almost surely.

Moreover, let there be an $N \times N$ matrix $\boldsymbol{H}$ such that the family

$$\left(\left\{ \boldsymbol{X}_1, \boldsymbol{X}_1^{-1}, \boldsymbol{X}_2, \boldsymbol{X}_2^{-1}, \dots \right\}, \{\boldsymbol{H}\} \right)$$

is almost surely asymptotically free, as $N \to \infty$.

Then, the family

$$\left(\left\{ \boldsymbol{X}_1 \boldsymbol{H} \boldsymbol{X}_1^{-1} \right\}, \left\{ \boldsymbol{X}_2 \boldsymbol{H} \boldsymbol{X}_2^{-1} \right\}, \dots \right)$$

is almost surely asymptotically free, too, as $N \to \infty$.

Chapter 6:

Compressive Sensing

Main Idea

Definition 12 *A source signal is sampled (sensed) in a compressive manner, if the sampling rate is below the rate required by the sampling theorem.*

In order to avoid aliasing and/or achieve a high quality reconstruction of the sampled (sensed) signal, the sampled signal is post-processed utilizing various forms of redundancy in the source signal.

Goal: Taking fewer samples of large data sets, e.g. 3D images in magnetic resonance tomography, speeds up the imaging process.

Vector Notation

Consider a source vector $\boldsymbol{x} \in \mathbb{R}^{K\times 1}$ and a sample vector $\boldsymbol{y} \in \mathbb{R}^{N\times 1}$. For a sensing matrix $\boldsymbol{S} \in \mathbb{R}^{N\times K}$ and additive noise $\boldsymbol{n} \in \mathbb{R}^{N\times 1}$, we get

$$\boldsymbol{y} = \boldsymbol{S}\boldsymbol{x} + \boldsymbol{n}$$

with $K > N$.

This is like overloaded CDMA.

Reconstruction by maximum a-posteriori detection

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}}\, \mathrm{p}_{\boldsymbol{x}|\boldsymbol{y}}(\boldsymbol{y}, \boldsymbol{x}) = \frac{\underset{\boldsymbol{x}}{\operatorname{argmin}}\, \mathrm{p}_{\boldsymbol{y}|\boldsymbol{x}}(\boldsymbol{y}, \boldsymbol{x})\mathrm{p}_{\boldsymbol{x}}(\boldsymbol{x})}{\int \mathrm{p}_{\boldsymbol{y}|\boldsymbol{x}}(\boldsymbol{y}, \tilde{\boldsymbol{x}})\mathrm{p}_{\boldsymbol{x}}(\tilde{\boldsymbol{x}})\mathrm{d}\tilde{\boldsymbol{x}}}$$

minimizes the probability that a wrong reconstruction is chosen.

Reconstruction Based on Minimum Distortion

For certain applications, it can be more important to minimize a certain distortion measure $\mathrm{d}(\cdot, \cdot\cdot)$ than to maximize the probability of correct detection

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{\xi}}{\operatorname{argmin}} \, \mathrm{d}(\boldsymbol{x}, \boldsymbol{\xi}) - \sum_i \mu_i f_i(\boldsymbol{\xi})$$

with $f_i(\cdot)$ and μ_i denoting side constraints modeling the redundancy of the source $\boldsymbol{x}$ and their respective Lagrange multipliers.

Common distortion measures:

- Mean-squared distortion $||\boldsymbol{x} - \hat{\boldsymbol{x}}||_2^2$
- Peak distortion $||\boldsymbol{x} - \hat{\boldsymbol{x}}||_\infty$

Common side constraints:

- Zero norm $f(\xi) = ||\xi||_0$
- One norm $f(\xi) = ||\xi||_1$ to reduce complexity of the zero norm

Chapter 7:

Vector Precoding

The Gaussian Vector Channel

Let the received vector be given by

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{t} + \boldsymbol{n}$$

where

- $\boldsymbol{t}$ is the transmitted vector
- $\boldsymbol{n}$ is uncorrelated (white) Gaussian noise
- $\boldsymbol{H}$ is a coupling matrix accounting for crosstalk

Crosstalk can be processed either at receiver or transmitter

Decorrelation at Transmitter

If the transmitter is a base-station and the receiver is a hand-held device, processing at the transmitter is preferred.

In a broadcast situation, processing at the transmitter is mandatory.

E.g. let the transmitted vector be

$$\boldsymbol{t} = \boldsymbol{H}^\dagger (\boldsymbol{H}\boldsymbol{H}^\dagger)^{-1} \boldsymbol{x}$$

where $\boldsymbol{x} = \boldsymbol{s}$ is the data to be sent.

Then,

$$\boldsymbol{y} = \boldsymbol{s} + \boldsymbol{n}.$$

No crosstalk anymore due to channel inversion.

Problems of Simple Channel Inversion

Channel inversion implies a significant power amplification, i.e.

$$\mathrm{E}\, \boldsymbol{t}^\dagger \boldsymbol{t} = \mathrm{E}\, \boldsymbol{x}^\dagger \left(\boldsymbol{H}\boldsymbol{H}^\dagger\right)^{-1} \boldsymbol{x} > \mathrm{E}\, \boldsymbol{x}^\dagger \boldsymbol{x}.$$

In particular, let

- $\alpha = \frac{K}{N} \leq 1$;
- the entries of $\boldsymbol{H}$ are i.i.d. with variance $1/N$.

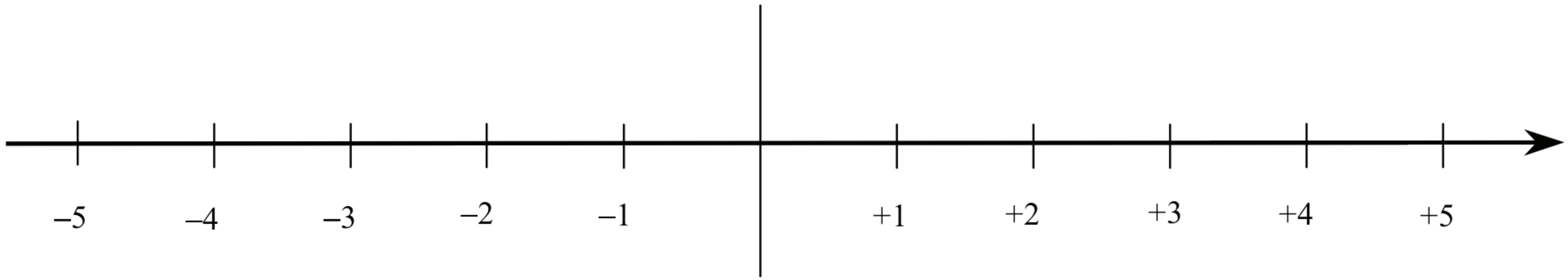
Then, for fixed aspect ratio α

$$\lim_{K\to\infty} \frac{\boldsymbol{x}^\dagger \left(\boldsymbol{H}\boldsymbol{H}^\dagger\right)^{-1} \boldsymbol{x}}{\boldsymbol{x}^\dagger \boldsymbol{x}} = \frac{1}{1-\alpha}$$

with probability 1.

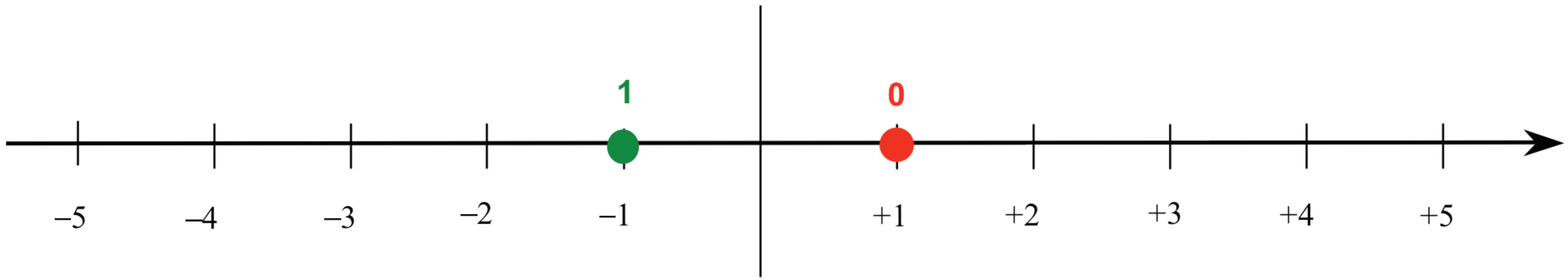
Tomlinson-Harashima Precoding

Tomlinson '71, Harashima & Miyakawa '72



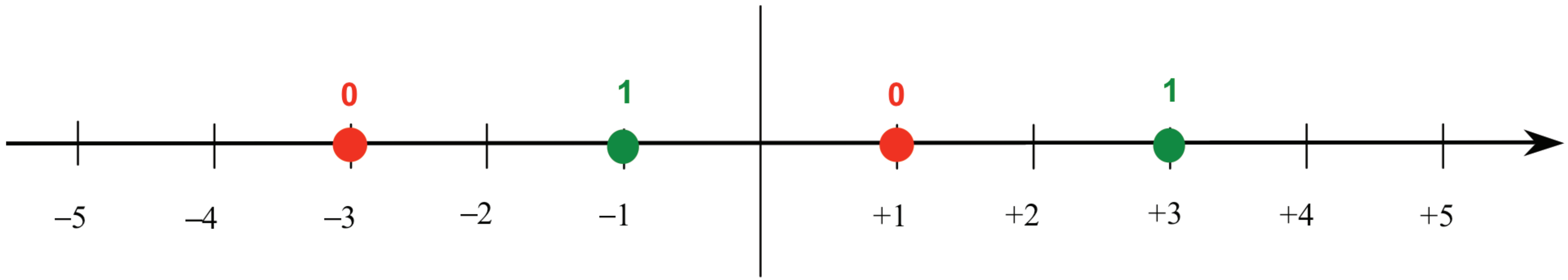
Tomlinson-Harashima Precoding

Tomlinson '71, Harashima & Miyakawa '72



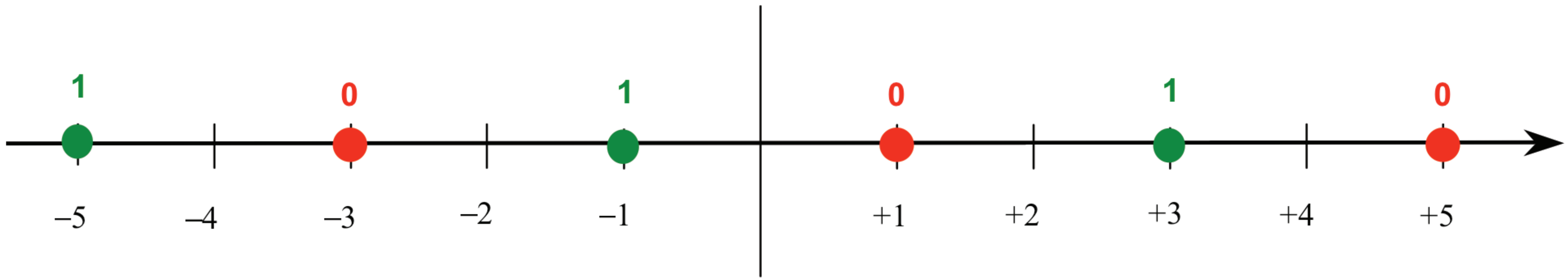
Tomlinson-Harashima Precoding

Tomlinson ’71; Harashima & Miyakawa ’72



Tomlinson-Harashima Precoding

Tomlinson '71, Harashima & Miyakawa '72



Instead of representing the logical "0" by $+1$, we present it by any element of the set $\{\ldots, -7, -3, +1, +5, \ldots\} = 4\mathbb{Z} + 1$. Correspondingly, the logical "1" is represented by any element of the set $4\mathbb{Z} - 1$.

Choose that representation that gives the smallest transmit power.

Generalized TH Precoding

Let $\mathcal{B}_0$ and $\mathcal{B}_1$ denote the sets presenting 0 and 1, resp.

Let $(s_1, s_2, s_3, \ldots, s_K) \in \{0, 1\}^K$ denote the data to be transmitted.

Then, the transmitted energy per data symbol is given by

$$E = \frac{1}{K} \min_{\boldsymbol{x} \in \mathcal{X}} \boldsymbol{x}^\dagger \boldsymbol{J} \boldsymbol{x}$$

with

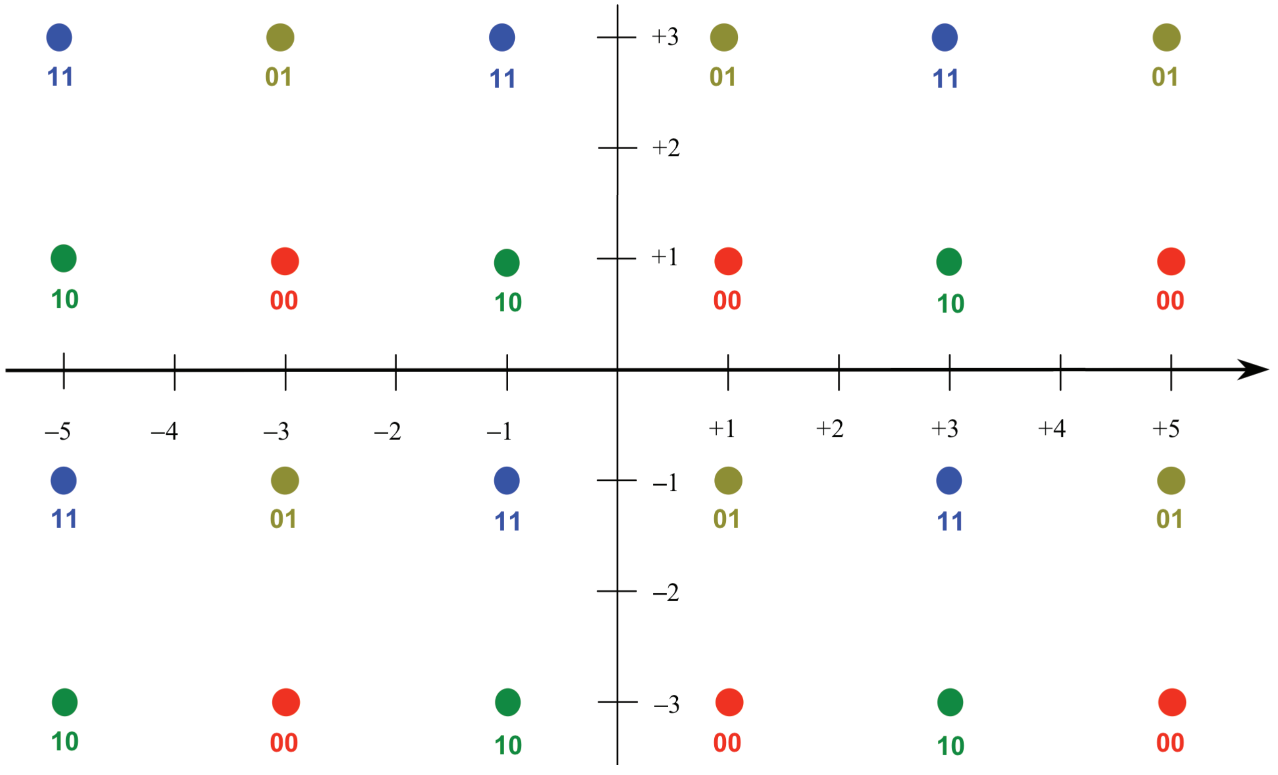
$$\mathcal{X} = \mathcal{B}_{s_1} \times \mathcal{B}_{s_2} \times \cdots \times \mathcal{B}_{s_K}$$

and

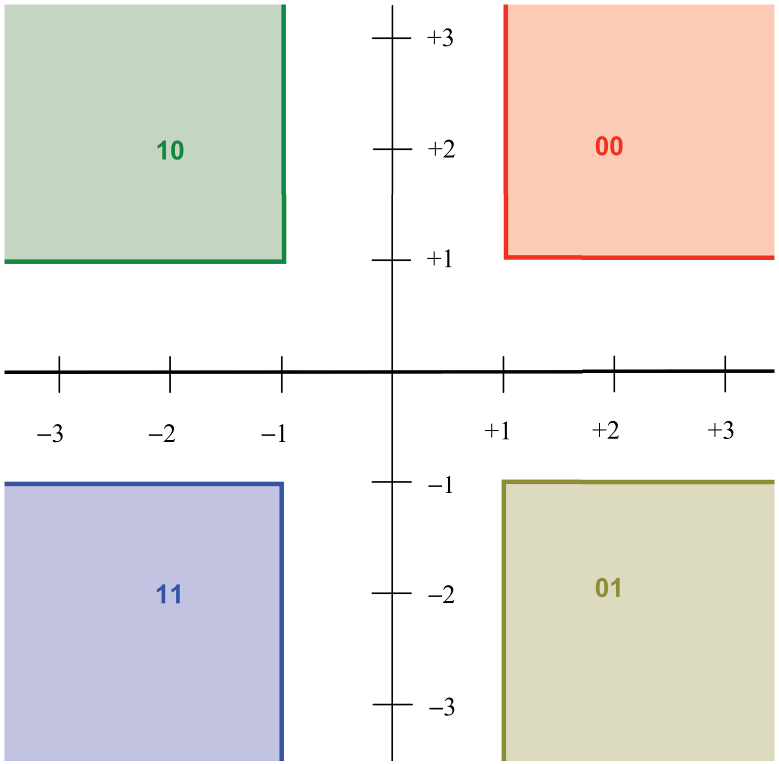
$$\boldsymbol{J} = (\boldsymbol{H}\boldsymbol{H}^\dagger)^{-1}.$$

What is a smart choice for $\mathcal{B}_0$ and $\mathcal{B}_1$?

Odd Integer Quadrature Lattice



Complex Convex Relaxation



. . . allows for convex programming.

Inverting Singular Channels

What happens if the channel is rank-deficient, e.g. $K > N$?

Can we precode without interference?

Inverting Singular Channels

What happens if the channel is rank-deficient, e.g. $K > N$?

Can we precode without interference?

The precoder produces

$$\lim_{\epsilon\to 0} \operatorname*{argmin}_{\boldsymbol{x}\in\mathcal{X}} \frac{\boldsymbol{x}^\dagger(\boldsymbol{H}\boldsymbol{H}^\dagger + \epsilon\mathbf{I})^{-1}\boldsymbol{x}}{K}$$

The received signal becomes

$$\boldsymbol{y} = \lim_{\epsilon\to 0} \boldsymbol{H}\boldsymbol{H}^\dagger(\boldsymbol{H}\boldsymbol{H}^\dagger + \epsilon\mathbf{I})^{-1}\boldsymbol{x} + \boldsymbol{n}.$$

Inverting Singular Channels

What happens if the channel is rank-deficient, e.g. $K > N$?

Can we precode without interference?

The precoder produces

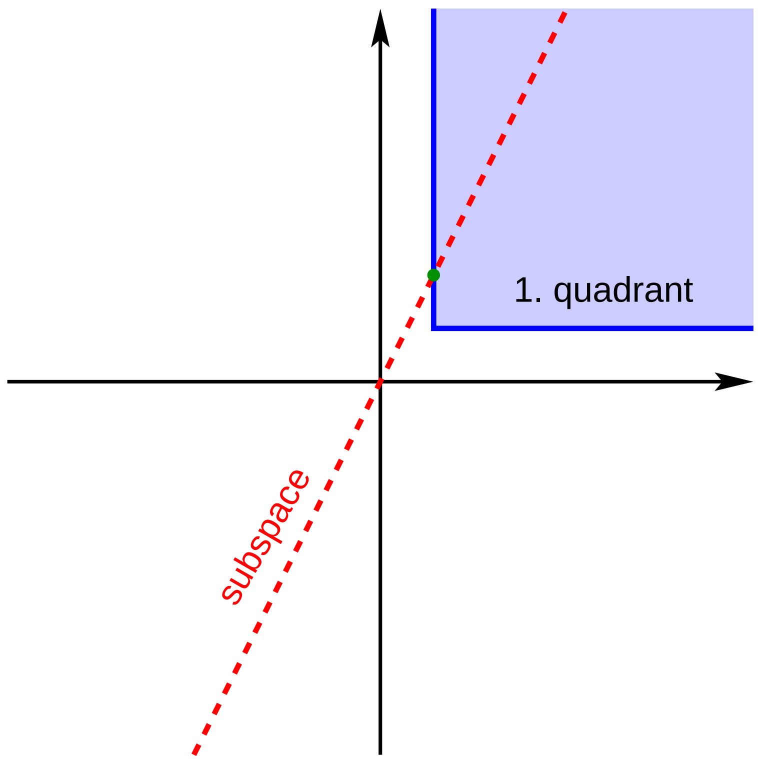
$$\lim_{\epsilon \to 0} \operatorname*{argmin}_{\boldsymbol{x} \in \mathcal{X}} \frac{\boldsymbol{x}^\dagger (\boldsymbol{H}\boldsymbol{H}^\dagger + \epsilon \mathbf{I})^{-1} \boldsymbol{x}}{K}$$

The received signal becomes

$$\boldsymbol{y} = \lim_{\epsilon \to 0} \boldsymbol{H}\boldsymbol{H}^\dagger (\boldsymbol{H}\boldsymbol{H}^\dagger + \epsilon \mathbf{I})^{-1} \boldsymbol{x} + \boldsymbol{n}.$$

If the energy is finite, there is no interference.

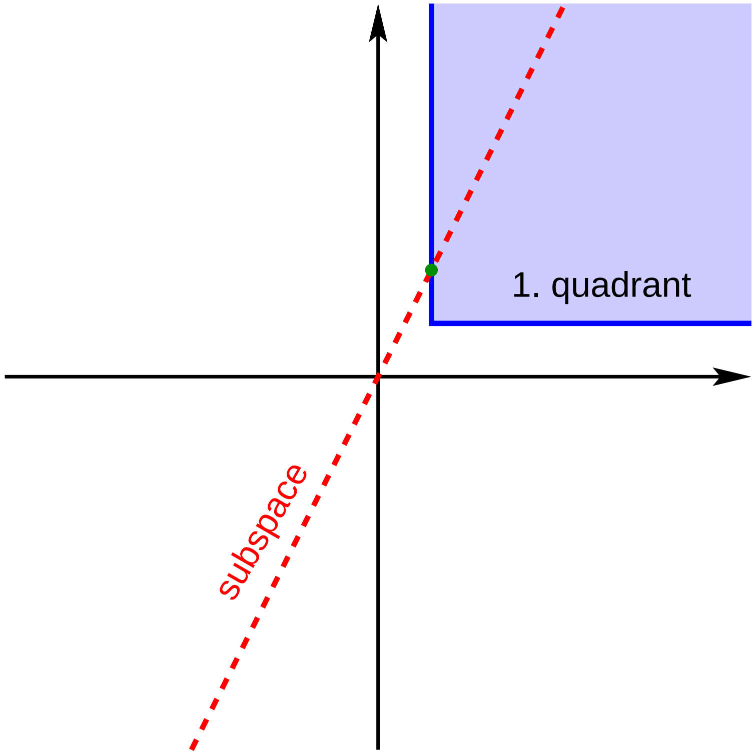
Overloaded Convex Precoding



The probability that a random N dimensional subspace in K real dimensions intersects the 1. K-tant is

$$P(K,N) = 2^{1-K} \sum_{\ell=0}^{N-1} \binom{K-1}{\ell}$$

Overloaded Convex Precoding



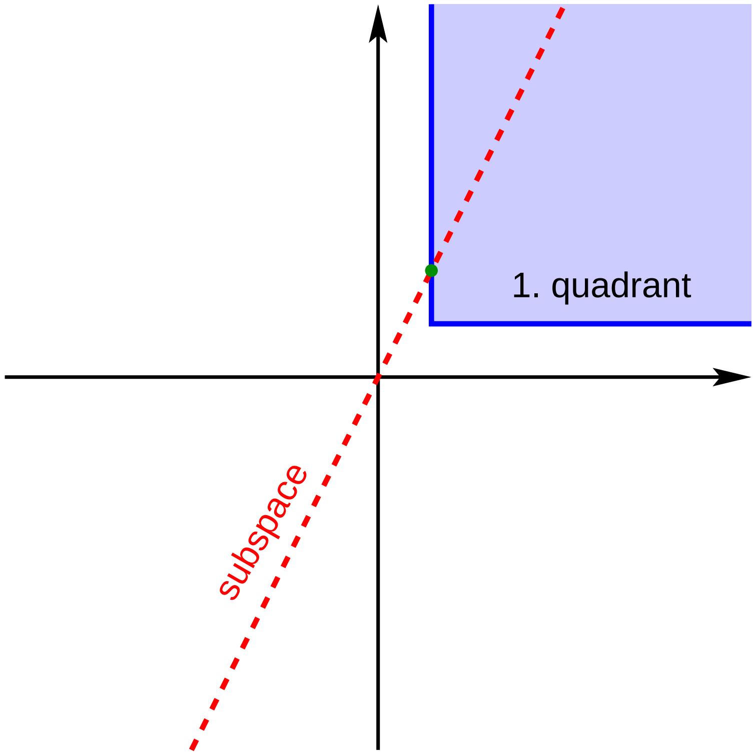
The probability that a random N dimensional subspace in K real dimensions intersects the 1. K-tant is

$$P(K,N) = 2^{1-K} \sum_{\ell=0}^{N-1} \binom{K-1}{\ell}$$

As K, N to infinity, we get

$$P(K,N) = \begin{cases} 1 & K < 2N \\ 1/2 & K = 2N \\ 0 & K > 2N \end{cases}$$

Overloaded Convex Precoding



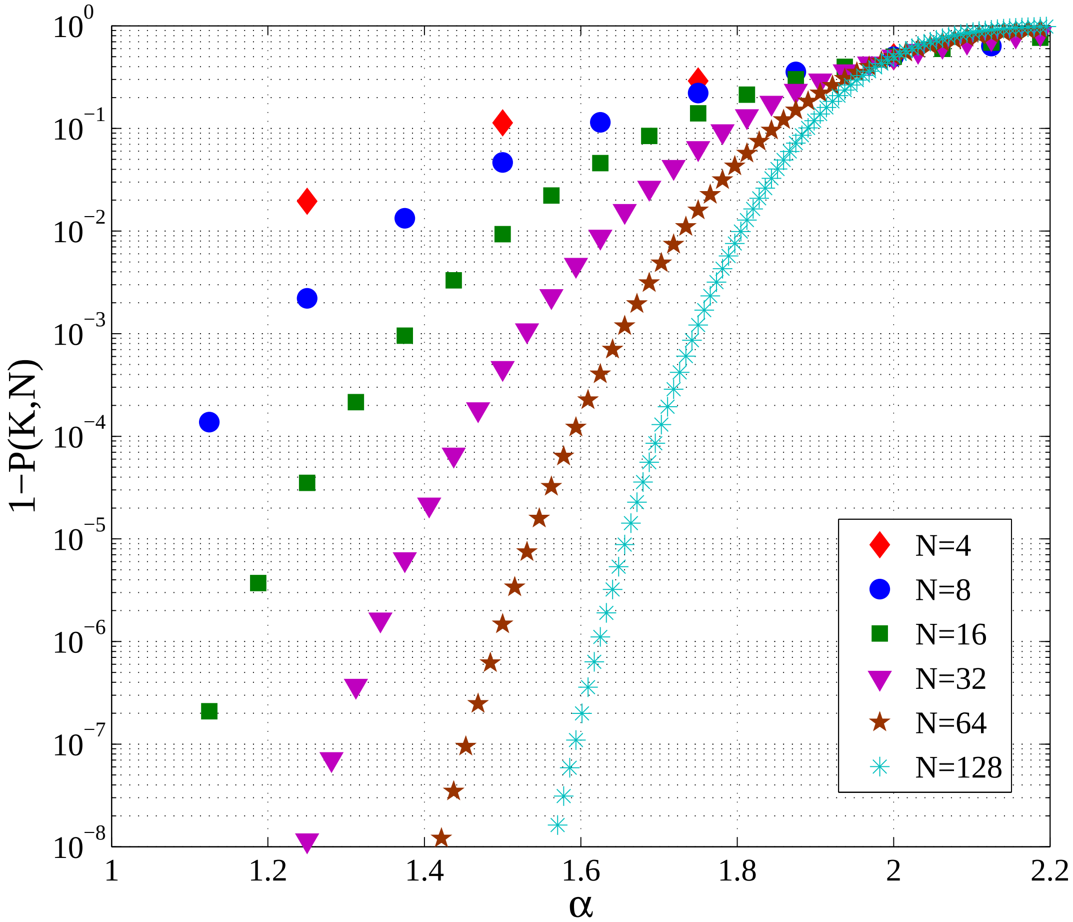
The probability that a random N dimensional subspace in K complex dimensions intersects the 1. K-tant is

$$P(K,N) = 2^{1-2K} \sum_{\ell=0}^{2N-1} \binom{2K-1}{\ell}$$

As K, N to infinity, we get

$$P(K,N) = \begin{cases} 1 & K < 2N \\ 1/2 & K = 2N \\ 0 & K > 2N \end{cases}$$

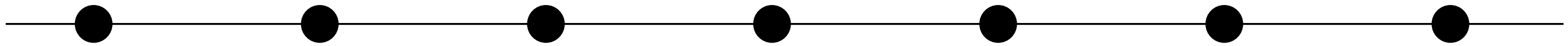
Overloaded Convex Precoding



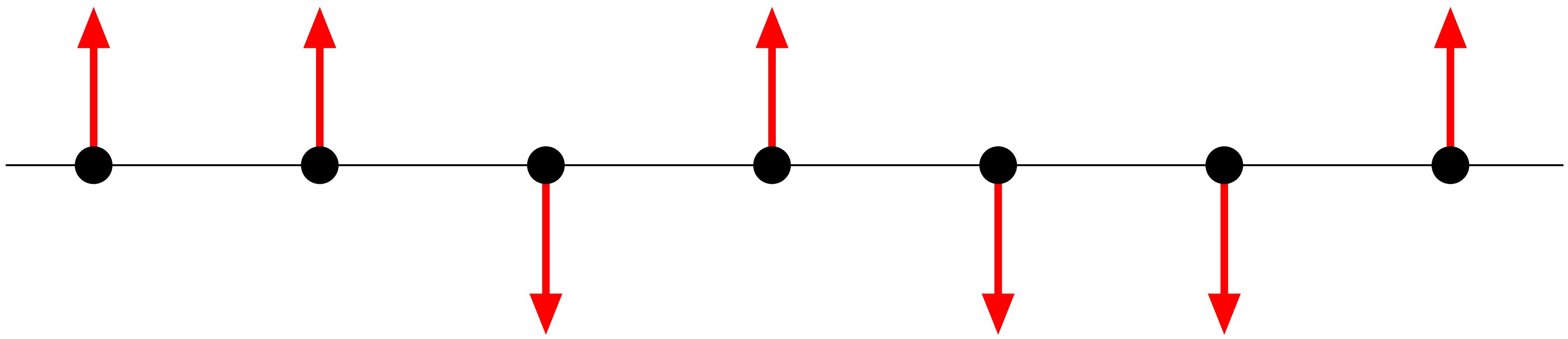
Chapter 8:

The Replica Method

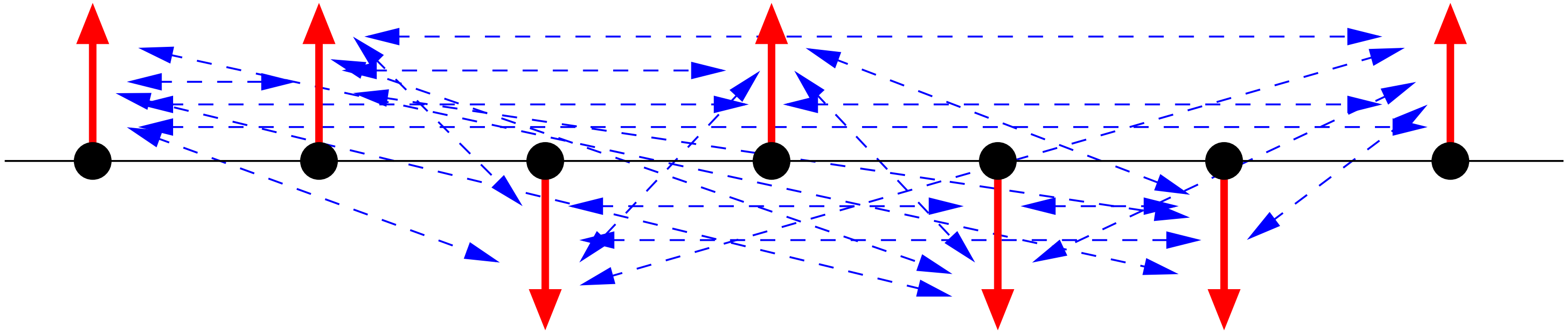
Spin Glasses



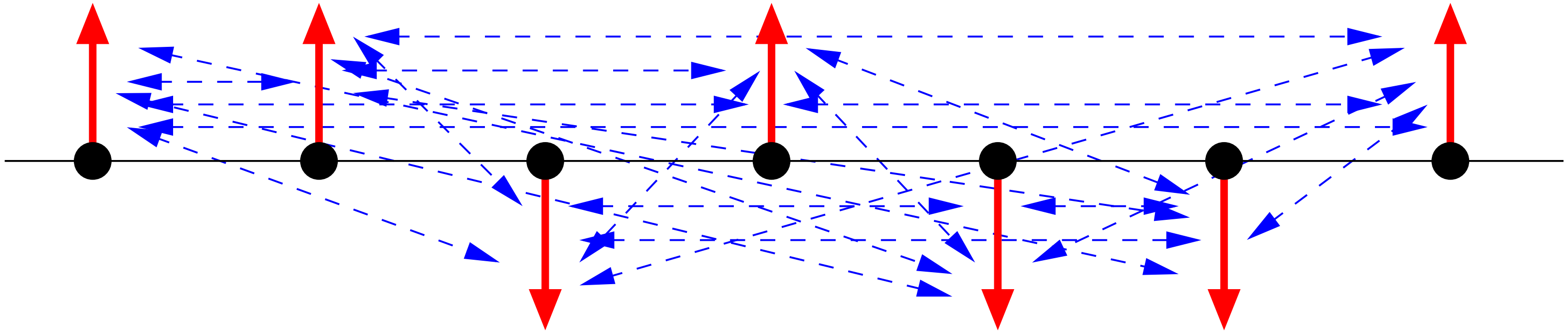
Spin Glasses



Spin Glasses



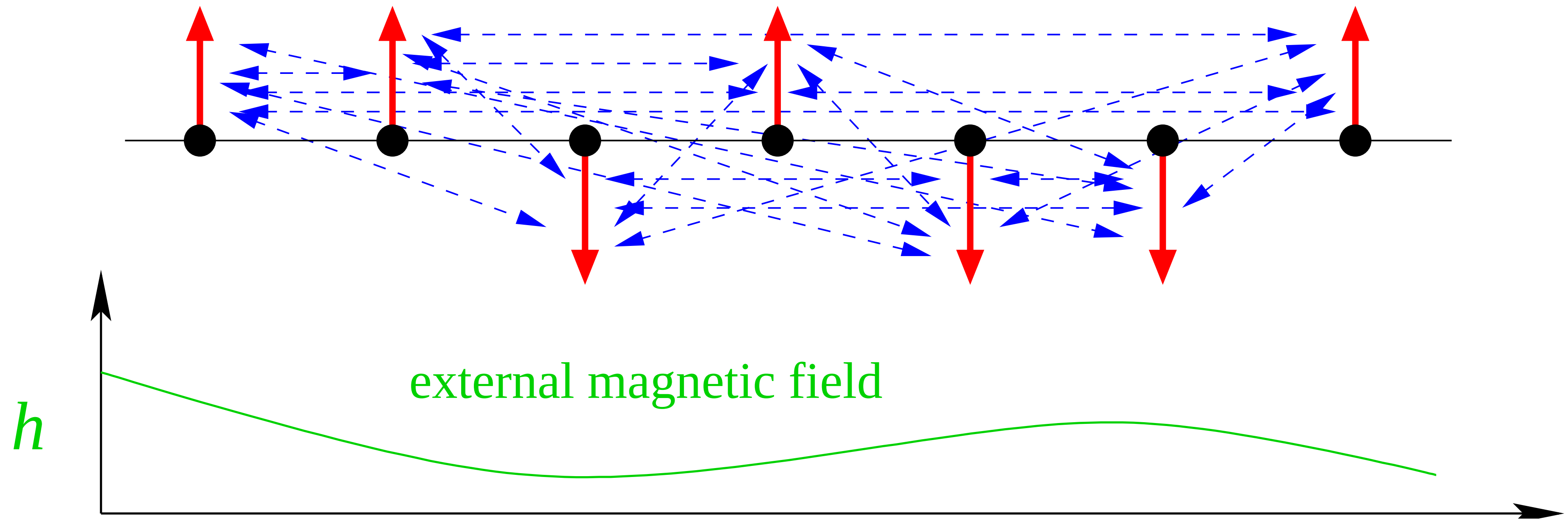
Spin Glasses



Energy function (Hamiltonian):

$$-\sum_i \sum_{j<i} r_{ij} x_i x_j$$

Spin Glasses



Energy function (Hamiltonian):

$$-\sum_i \sum_{j<i} r_{ij} x_i x_j - \sum_i h_i x_i$$

Optimal Detection of Vector Channel

$$\boldsymbol{y} = \boldsymbol{S}\boldsymbol{x} + \boldsymbol{n}$$

Optimal Detection of Vector Channel

$$\boldsymbol{y} = \boldsymbol{S}\boldsymbol{x} + \boldsymbol{n}$$

Best estimate for transmitted data:

$$\hat{\boldsymbol{x}} = \operatorname*{argmin}_{\boldsymbol{x} \in \{\pm 1\}^K} ||\boldsymbol{y} - \boldsymbol{S}\boldsymbol{x}||$$

Optimal Detection of Vector Channel

$$\boldsymbol{y} = \boldsymbol{S}\boldsymbol{x} + \boldsymbol{n}$$

Best estimate for transmitted data:

$$\begin{aligned} \hat{\boldsymbol{x}} &= \operatorname*{argmin}_{\boldsymbol{x}\in\{\pm 1\}^K} ||\boldsymbol{y} - \boldsymbol{S}\boldsymbol{x}|| \\ &= \operatorname*{argmin}_{\boldsymbol{x}\in\{\pm 1\}^K} -\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{x} - \boldsymbol{h}^{\mathrm{T}}\boldsymbol{x} + \boldsymbol{y}^{\mathrm{T}}\boldsymbol{y} \end{aligned} \qquad \text{with} \qquad \begin{aligned} \boldsymbol{R} &= -2\boldsymbol{S}^{\mathrm{T}}\boldsymbol{S} \\ \boldsymbol{h} &= 2\boldsymbol{S}^{\mathrm{T}}\boldsymbol{y} \end{aligned}$$

Optimal Detection of Vector Channel

$$\boldsymbol{y} = \boldsymbol{S}\boldsymbol{x} + \boldsymbol{n}$$

Best estimate for transmitted data:

$$\begin{aligned}
\hat{\boldsymbol{x}} &= \operatorname*{argmin}_{\boldsymbol{x}\in\{\pm 1\}^K} ||\boldsymbol{y} - \boldsymbol{S}\boldsymbol{x}|| \\
&= \operatorname*{argmin}_{\boldsymbol{x}\in\{\pm 1\}^K} -\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{x} - \boldsymbol{h}^{\mathrm{T}}\boldsymbol{x} \qquad \text{with} \qquad \begin{aligned} \boldsymbol{R} &= -2\boldsymbol{S}^{\mathrm{T}}\boldsymbol{S} \\ \boldsymbol{h} &= 2\boldsymbol{S}^{\mathrm{T}}\boldsymbol{y} \end{aligned} \\
&= \operatorname*{argmin}_{\boldsymbol{x}\in\{\pm 1\}^K} -\sum_i \sum_{j<i} r_{ij}x_i x_j - \sum_i h_i x_i - \frac{1}{2}\sum_i r_{ii}x_i^2
\end{aligned}$$

Optimal Detection of Vector Channel

$$\boldsymbol{y} = \boldsymbol{S}\boldsymbol{x} + \boldsymbol{n}$$

Best estimate for transmitted data:

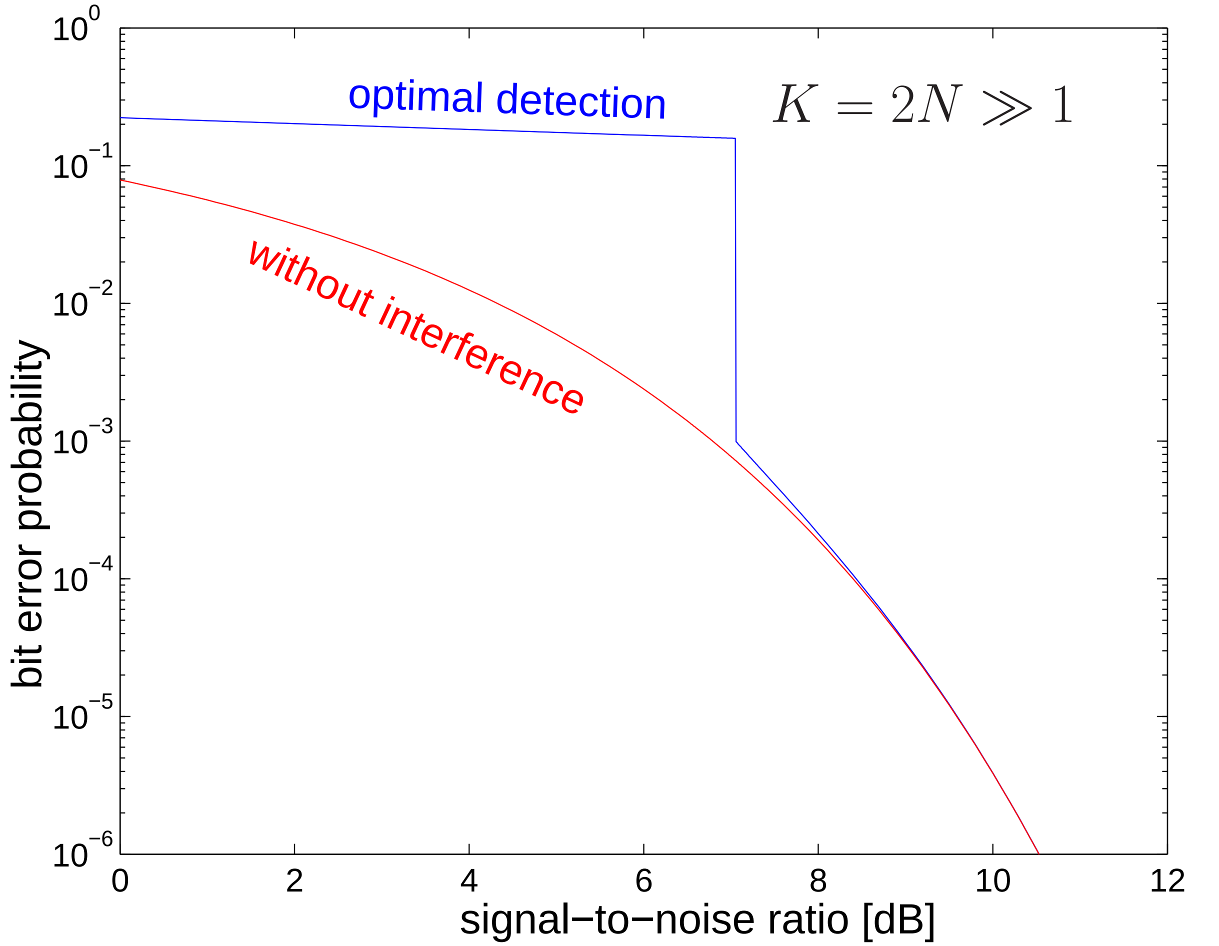
$$\begin{aligned}
\hat{\boldsymbol{x}} &= \underset{\boldsymbol{x}\in\{\pm 1\}^K}{\operatorname{argmin}} ||\boldsymbol{y} - \boldsymbol{S}\boldsymbol{x}|| \\
&= \underset{\boldsymbol{x}\in\{\pm 1\}^K}{\operatorname{argmin}} -\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{x} - \boldsymbol{h}^{\mathrm{T}}\boldsymbol{x} \\
&= \underset{\boldsymbol{x}\in\{\pm 1\}^K}{\operatorname{argmin}} -\sum_i \sum_{j<i} r_{ij} x_i x_j - \sum_i h_i x_i
\end{aligned}$$

with

$$\begin{aligned}
\boldsymbol{R} &= -2\boldsymbol{S}^{\mathrm{T}}\boldsymbol{S} \\
\boldsymbol{h} &= 2\boldsymbol{S}^{\mathrm{T}}\boldsymbol{y}
\end{aligned}$$

Minimization of the energy function of a spin glass!

A Phase Transition in Random CDMA



Energy vs. Entropy

The following two tasks are dual:

- Minimize the energy for fixed entropy
- Maximize the entropy for fixed energy

Consider free energy

$$F(X) = E(X) - TH(X)$$

and read the temperature (or its inverse) as Lagrange multiplier.

For the dual problem have

$$-\frac{1}{T}F(X) = H(X) - \frac{1}{T}E(X)$$

The Meaning of the Energy Function

In physics, the energy function varies with the force causing the potential.

Theoretically speaking, the choice of the energy function is arbitrary as long as it is uniformly bounded from below.

Nature maximizes entropy for a given energy.

In communications engineering, the energy function is the metric used by the decoder.

The decoder does the dual job of nature, to minimize the metric for a given output entropy.

Since the decoder dictates the thermodynamics of our toy universe, the same holds true if the decoder uses a suboptimal (wrong) or insufficient metric, perhaps due to lack of knowledge about the channel state.

The free choice of the energy function allows to analyze mismatched receivers.

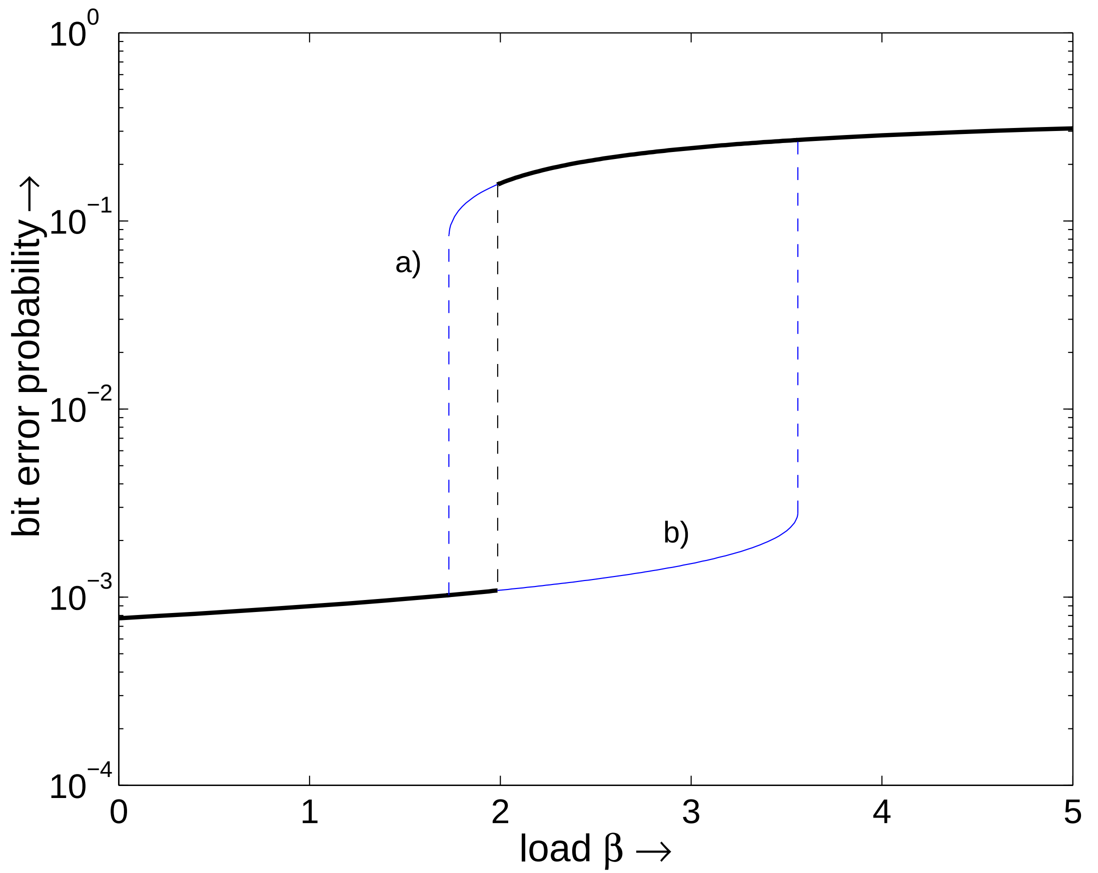
LMMSE Detector with Mismatched Powers

Theorem 22 *Let* $(U_1, \ldots, U_K)$ *be an arbitrary sequence of non-negative numbers such that, as* $K \to \infty$*, the empirical joint cdf of the pairs* $\{(U_k, P_k) : k = 1, \ldots, K\}$ *converges weakly to a given non-random cdf* $F(u,p)$*. Moreover, let the* P_k*s be uniformly bounded from above and the* U_k*s uniformly bounded from below by a positive number for all* K*. Then, the multiuser efficiency of the mismatched LMMSE detector assuming powers* $\{U_k\}$ *instead of the true powers* $\{P_k\}$ *in the standard random spreading model converges as* $K = \alpha N \to \infty$ *almost surely to*

$$\eta \frac{1 + \alpha \displaystyle\int \frac{u}{(\sigma^2 + u\eta)^2}\, dF(u,p)}{1 + \alpha \displaystyle\int \frac{p}{(\sigma^2 + u\eta)^2}\, dF(u,p)} \quad \text{where} \quad \eta = \left(1 + \alpha \int \frac{u}{\sigma^2 + u\eta}\, dF(u,p)\right)^{-1}$$

is the multiuser efficiency of an LMMSE detector of a "virtual channel" having powers given by $\{U_k\}$ *instead of* $\{P_k\}$.

Phase Transitions and Neural Networks



Individually Optimum ML Detector

Let $\mathcal{A} = \{+1; -1\}$, the chips of any user be i.i.d. random variables with finite variance and vanishing odd moments, the powers of all users identical, and $N, K \to \infty$, but $\alpha = K/N$ fixed. Then, the multiuser efficiency is a solution to the fixed point equation

$$\frac{1}{\eta_{\mathrm{IO}}} = 1 + \frac{\alpha}{\sigma^2}\left[1 - \sqrt{\frac{\eta_{\mathrm{IO}}}{2\pi\sigma^2}} \int_{\mathbb{R}} \tanh\left(\frac{\eta_{\mathrm{IO}}}{\sigma^2}x\right) \exp\left(-\frac{\eta_{\mathrm{IO}}(x-1)^2}{2\sigma^2}\right) \mathrm{d}x\right].$$

In case the fixed point equation has multiple solutions, the correct one is that solution for which the term

$$\frac{\eta_{\mathrm{IO}}}{\sigma^2} + \frac{\eta_{\mathrm{IO}} - \log \eta_{\mathrm{IO}}}{2\alpha} - \sqrt{\frac{\eta_{\mathrm{IO}}}{2\pi\sigma^2}} \int_{\mathbb{R}} \log\left[\cosh\left(\frac{\eta_{\mathrm{IO}}}{\sigma^2}x\right)\right] \exp\left(-\frac{\eta_{\mathrm{IO}}(x-1)^2}{2\sigma^2}\right) \mathrm{d}x$$

is smallest.

Chapter 9:

Examples for Replica Calculations

Bit Error Rate for Large CDMA

Consider the analysis of an asymptotically large CDMA systems with arbitrary joint distribution of the variances of the random chips.

The vector-valued, real additive white Gaussian noise channel is characterized by the conditional pdf

$$\mathrm{p}_{\boldsymbol{y}|\boldsymbol{x},\boldsymbol{H}}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{H}) = \frac{\mathrm{e}^{-\frac{1}{2\sigma_0^2}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})}}{(2\pi\sigma_0^2)^{\frac{N}{2}}}.$$

Let the detector be characterized by the assumed conditional probability distribution

$$\breve{\mathrm{p}}_{\boldsymbol{y}|\boldsymbol{x},\boldsymbol{H}}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{H}) = \frac{\mathrm{e}^{-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})}}{(2\pi\sigma^2)^{\frac{N}{2}}}$$

and the assumed prior distribution $\breve{\mathrm{p}}_{\boldsymbol{x}}(\boldsymbol{x})$.

Energy Function

Applying Bayes' law, we find

$$\breve{\mathrm{p}}_{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{H}}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{H}) = \frac{\mathrm{e}^{-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})+\log\breve{\mathrm{p}}_{\boldsymbol{x}}(\boldsymbol{x})}}{\int \mathrm{e}^{-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})}\,\mathrm{d}\breve{\mathrm{P}}_{\boldsymbol{x}}(\boldsymbol{x})}.$$

Since the Boltzmann distribution holds for any temperature T, we set w.l.o.g. $T = 1$ and find the appropriate energy function to be

$$||\boldsymbol{x}|| = \frac{1}{2\sigma^2}\left(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}\right)^{\mathrm{T}}\left(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}\right) - \log\breve{\mathrm{p}}_{\boldsymbol{x}}(\boldsymbol{x})\,.$$

This choice of the energy function ensures that the thermodynamic equilibrium models the detector defined by the assumed conditional and prior distributions.

Free Energy

We plug the energy function into the free energy per user and average over the channel and the true joint distribution of $\boldsymbol{x}$ and $\boldsymbol{y}$

$$\frac{\mathrm{F}(\boldsymbol{x})}{K} = -\frac{1}{K}\mathop{\mathrm{E}}_{\boldsymbol{H}}\int\int_{\mathbb{R}^N}\log\left(\int \mathrm{e}^{-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})}\mathrm{d}\breve{\mathrm{P}}_{\boldsymbol{x}}(\boldsymbol{x})\right)\frac{\mathrm{e}^{-\frac{1}{2\sigma_0^2}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})}}{(2\pi\sigma_0^2)^{\frac{N}{2}}}\mathrm{d}\boldsymbol{y}\mathrm{d}\mathrm{P}_{\boldsymbol{x}}(\boldsymbol{x})$$

$$= -\frac{1}{K}\lim_{n\to 0}\frac{\partial}{\partial n}\log\mathop{\mathrm{E}}_{\boldsymbol{H}}\int\int_{\mathbb{R}^N}\left(\int \mathrm{e}^{-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})}\mathrm{d}\breve{\mathrm{P}}_{\boldsymbol{x}}(\boldsymbol{x})\right)^n\frac{\mathrm{e}^{-\frac{1}{2\sigma_0^2}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})}}{(2\pi\sigma_0^2)^{\frac{N}{2}}}\mathrm{d}\boldsymbol{y}\mathrm{d}\mathrm{P}_{\boldsymbol{x}}(\boldsymbol{x})$$

$$= -\frac{1}{K}\lim_{n\to 0}\frac{\partial}{\partial n}\log\underbrace{\int\frac{\int_{\mathbb{R}^N}\mathop{\mathrm{E}}_{\boldsymbol{H}}\prod_{a=0}^{n}\mathrm{e}^{-\frac{1}{2\sigma_a^2}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}_a)^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}_a)}\mathrm{d}\boldsymbol{y}}{(2\pi\sigma_0^2)^{\frac{N}{2}}}\prod_{a=0}^{n}\mathrm{d}\mathrm{P}_a(\boldsymbol{x}_a)}_{\triangleq\,\Xi_n}$$

with $\sigma_a = \sigma, \forall a \geq 1$, $\mathrm{P}_0(\boldsymbol{x}) = \mathrm{P}_{\boldsymbol{x}}(\boldsymbol{x})$, and $\mathrm{P}_a(\boldsymbol{x}) = \breve{\mathrm{P}}_{\boldsymbol{x}}(\boldsymbol{x}), \forall a \geq 1$.

Quenched Random Variables

The argument of the logarithm is given by

$$\Xi_n = \int \prod_{c=1}^{N} \frac{\int_{\mathbb{R}} \mathop{\mathrm{E}}_{\boldsymbol{H}} \prod_{a=0}^{n} \exp\left[-\frac{1}{2\sigma_a^2}\left(y_c - \sum_{k=1}^{K} h_{ck} x_{ak}\right)^2\right] \mathrm{d}y_c}{\sqrt{2\pi}\sigma_0} \prod_{a=0}^{n} \mathrm{dP}_a(\boldsymbol{x}_a)\,,$$

The integrand depends on $\boldsymbol{x}_a$ only through

$$v_{ac} \overset{\triangle}{=} \frac{1}{\sqrt{\alpha}} \sum_{k=1}^{K} h_{ck} x_{ak}, \qquad a = 0, \ldots, n.$$

For $\mathrm{E}\, h_{ck}^* h_{c'k'} = \frac{1}{N} w_{ck}^2 \delta_{cc'} \delta_{kk'}$, these quantities v_{ac} can be regarded, in the limit $K \to \infty$ as jointly Gaussian random variables with zero mean and covariances

$$Q_{ab}[c] = \mathop{\mathrm{E}}_{\boldsymbol{H}} v_{ac} v_{bc} = \frac{1}{K} \boldsymbol{x}_a \overset{(c)}{\bullet} \boldsymbol{x}_b$$

where the parametric inner products are defined by $\boldsymbol{x}_a \overset{(c)}{\bullet} \boldsymbol{x}_b \overset{\triangle}{=} \sum_{k=1}^{K} x_{ak} x_{bk} w_{ck}^2.$

Change of Variables

To perform integration, the $K(n+1)$-dimensional space spanned by the replicas and the vector $\boldsymbol{x}_0$ is split into subshells

$$S(Q_{ab}[c]) \triangleq \left\{ \boldsymbol{x}_0, \dots, \boldsymbol{x}_n \middle| \boldsymbol{x}_a \overset{(c)}{\bullet} \boldsymbol{x}_b = KQ_{ab}[c] \right\}$$

where the inner product of two different vectors $\boldsymbol{x}_a$ and $\boldsymbol{x}_b$ is constant.

The splitting of the $K(n+1)$-dimensional space is depending on the chip time c. With this splitting of the space, we find for $K \to \infty$

$$\Xi_n = \int\limits_{\mathbb{R}^{N(n+1)(n+2)/2}} \mathrm{e}^{K I(\boldsymbol{Q}[\cdot])} \prod_{c=1}^{N} \mathrm{e}^{G(\boldsymbol{Q}[c])} \prod_{a \le b} \mathrm{d}Q_{ab}[c],$$

with appropriate choices of the function $I(\boldsymbol{Q}[\cdot])$ and $G(\boldsymbol{Q}[c])$.

Change of Variables (cont'd)

Hereby,

$$e^{KI(\boldsymbol{Q}[\cdot])} = \int \left[\prod_{a\le b}\prod_{c=1}^{N} \delta\left(\frac{\boldsymbol{x}_a \overset{(c)}{\bullet} \boldsymbol{x}_b}{N} - \alpha Q_{ab}[c]\right)\right] \prod_{a=0}^{n} \mathrm{dP}_a(\boldsymbol{x}_a)$$

denotes the probability weight of the subshell and

$$e^{G(\boldsymbol{Q}[c])} = \frac{1}{\sqrt{2\pi}\sigma_0} \int_{\mathbb{R}} \underset{\boldsymbol{H}}{\mathrm{E}} \prod_{a=0}^{n} \exp\left[-\frac{\alpha}{2\sigma_a^2}\left(\frac{y_c}{\sqrt{\alpha}} - v_{ac}(\boldsymbol{Q}[c])\right)^2\right] \mathrm{d}y_c.$$

This procedure is a change of integration variables in multiple dimensions where the integration of an exponential function over the replicas has been replaced by integration over the quenched variables $Q_{ab}[\cdot]$. In the following the blue and green terms are evaluated subsequently.

Subshell Probability

We write the Dirac measure as the inverse Laplace transform of a constant

$$\delta\left(\frac{\boldsymbol{x}_a \overset{(c)}{\bullet} \boldsymbol{x}_b}{N} - \alpha Q_{ab}[c]\right) = \frac{1}{2\pi \mathrm{j}} \oint \exp\left[\tilde{Q}_{ab}[c]\left(\frac{\boldsymbol{x}_a \overset{(c)}{\bullet} \boldsymbol{x}_b}{N} - \alpha Q_{ab}[c]\right)\right] \mathrm{d}\tilde{Q}_{ab}[c].$$

Then, the measure $\mathrm{e}^{KI(\boldsymbol{Q}[\cdot])}$ can be expressed as

$$\begin{aligned}
\mathrm{e}^{KI(\boldsymbol{Q}[\cdot])} &= \int \left[\prod_{c=1}^{N}\prod_{a\le b} \oint \mathrm{e}^{\tilde{Q}_{ab}[c]\left(\frac{\boldsymbol{x}_a \overset{(c)}{\bullet} \boldsymbol{x}_b}{N} - \alpha Q_{ab}[c]\right)} \frac{\mathrm{d}\tilde{Q}_{ab}[c]}{2\pi \mathrm{j}}\right] \prod_{a=0}^{n} \mathrm{dP}_a(\boldsymbol{x}_a) \\
&= \oint \mathrm{e}^{-\alpha \sum\limits_{c=1}^{N}\sum\limits_{a\le b} \tilde{Q}_{ab}[c] Q_{ab}[c]} \left(\prod_{k=1}^{K} M_k\left(\tilde{\boldsymbol{Q}}[\cdot]\right)\right) \prod_{c=1}^{N}\prod_{a\le b} \frac{\mathrm{d}\tilde{Q}_{ab}[c]}{2\pi \mathrm{j}}
\end{aligned}$$

with the moment generating function

$$M_k\left(\tilde{\boldsymbol{Q}}[\cdot]\right) = \int \mathrm{e}^{\frac{1}{N}\sum\limits_{a\le b}\sum\limits_{c=1}^{N} \tilde{Q}_{ab}[c] x_{ak} x_{bk} w_{ck}^2} \prod_{a=0}^{n} \mathrm{dP}_a(x_{ak}).$$

Replica Symmetry

The integrals over $\boldsymbol{Q}[\cdot]$ and $\tilde{\boldsymbol{Q}}[\cdot]$ shall be solved by saddle point integration, but finding for the optimal matrices $\boldsymbol{Q}[\cdot]$ and $\tilde{\boldsymbol{Q}}[\cdot]$ seems a hopeless task.

So we make an ansatz and hope that it will work.

$$\boldsymbol{Q}[c] \triangleq \begin{bmatrix} p_{0c} & m_c & \cdots & m_c & m_c \\ m_c & p_c & q_c & \cdots & q_c \\ \vdots & q_c & \ddots & \ddots & \vdots \\ m_c & \vdots & \ddots & p_c & q_c \\ m_c & q_c & \cdots & q_c & p_c \end{bmatrix} \quad , \quad \tilde{\boldsymbol{Q}}[c] \triangleq \begin{bmatrix} \frac{G_{0c}}{2} & \frac{G_c}{2} & \cdots & \frac{G_c}{2} & \frac{G_c}{2} \\ \frac{G_c}{2} & E_c & F_c & \cdots & F_c \\ \vdots & F_c & \ddots & \ddots & \vdots \\ \frac{G_c}{2} & \vdots & \ddots & E_c & F_c \\ \frac{G_c}{2} & F_c & \cdots & F_c & E_c \end{bmatrix}$$

with some macroscopic parameters q_c, p_c, m_c, p_{0c} and $F_c, E_c, \frac{G_c}{2}, \frac{G_{0c}}{2}$.

We distinguish the cross-correlation between different replicas and the autocorrelation of an individual replica as well as between true and assumed variables.

Replica Symmetry *(cont'd)*

The assumption of replica symmetry leads to

$$\sum_{a \le b} \tilde{Q}_{ab}[c] Q_{ab}[c] = n E_c m_c + \frac{n(n-1)}{2} F_c q_c + \frac{G_{0c} p_{0c}}{2} + \frac{n}{2} G_c p_c$$

and

$$M_k\{E, F, G, G_0\} = \int\limits_{\mathbb{R}^{n+1}} \mathrm{e}^{\frac{\tilde{G}_{0k}}{2} x_{0k}^2 + \sum\limits_{a=1}^{n} \tilde{E}_k x_{0k} x_{ak} + \frac{\tilde{G}_k}{2} x_{ak}^2 + \sum\limits_{b=a+1}^{n} \tilde{F}_k x_{ak} x_{bk}} \prod_{a=0}^{n} \mathrm{dP}_a(x_{ak})$$

where

$$\tilde{E}_k \triangleq \frac{1}{N} \sum_{c=1}^{N} E_c w_{ck}^2 \quad , \qquad \tilde{F}_k \triangleq \frac{1}{N} \sum_{c=1}^{N} F_c w_{ck}^2$$

$$\tilde{G}_k \triangleq \frac{1}{N} \sum_{c=1}^{N} G_c w_{ck}^2 \quad , \qquad \tilde{G}_{0k} \triangleq \frac{1}{N} \sum_{c=1}^{N} G_{0c} w_{ck}^2 .$$

Hubbard Stratonovich Transform

We complete the square in $M_k\{E,F,G,G_0\}$ and apply the Hubbard-Stratonovich transform to linearize the exponential argument.

$$M_k\{E,F,G,G_0\} =$$

$$= \int \mathrm{e}^{\frac{\tilde G_{0k}}{2}x_{0k}^2+\sum\limits_{a=1}^{n}\tilde E_k x_{0k}x_{ak}+\frac{\tilde G_k}{2}x_{ak}^2+\sum\limits_{b=a+1}^{n}\tilde F_k x_{ak}x_{bk}} \prod_{a=0}^{n} \mathrm{dP}_a(x_{ak})$$

$$= \int \mathrm{e}^{\frac{\tilde G_{0k}}{2}x_{0k}^2+\frac{\tilde F_k}{2}\left(\sum\limits_{a=1}^{n}x_{ak}\right)^2+\sum\limits_{a=1}^{n}\tilde E_k x_{0k}x_{ak}+\frac{\tilde G_k-\tilde F_k}{2}x_{ak}^2} \prod_{a=0}^{n} \mathrm{dP}_a(x_{ak})$$

$$= \iint \mathrm{e}^{\frac{\tilde G_{0k}}{2}x_{0k}^2+\sum\limits_{a=1}^{n}\tilde E_k x_{0k}x_{ak}+\sqrt{\tilde F_k}z x_{ak}+\frac{\tilde G_k-\tilde F_k}{2}x_{ak}^2} \mathrm{D}z \prod_{a=0}^{n} \mathrm{dP}_a(x_{ak})$$

$$= \int \mathrm{e}^{\frac{\tilde G_{0k}}{2}x_k^2} \int\left(\int \mathrm{e}^{\tilde E_k x_k \breve x_k+\sqrt{\tilde F_k}z\breve x_k+\frac{\tilde G_k-\tilde F_k}{2}\breve x_k^2}\mathrm{d}\breve{\mathrm{P}}_{\breve x_k}(\breve x_k)\right)^n \mathrm{D}z\,\mathrm{dP}_{x_k}(x_k)$$

The $n+1$-dimensional integral over the prior distribution has become a 3D integral.

Green Replica Symmetry

We construct the variables v_{ac} out of independent zero-mean, unit-variance Gaussian random variables u_c, t_c, z_{ac} by

$$
\begin{aligned}
v_{0c} &= u_c\sqrt{p_{0c} - \frac{m_c^2}{q_c}} - t_c\frac{m_c}{\sqrt{q_c}} \\
v_{ac} &= z_{ac}\sqrt{p_c - q_c} - t_c\sqrt{q_c}, \quad a > 0
\end{aligned}
$$

and get

$$
\begin{aligned}
\mathrm{e}^{G(m_c,q_c,p_c,p_{0c})} &= \frac{1}{\sqrt{2\pi}\sigma_0}\int\limits_{\mathbb{R}^2}\int\limits_{\mathbb{R}} \exp\left[-\frac{\alpha}{2\sigma_0^2}\left(u_c\sqrt{p_{0c}-\frac{m_c^2}{q_c}} - \frac{t_c m_c}{\sqrt{q_c}} - \frac{y_c}{\sqrt{\alpha}}\right)^2\right]\mathrm{D}u_c \\
&\quad\times\left[\int\limits_{\mathbb{R}} \exp\left[-\frac{\alpha}{2\sigma^2}\left(z_c\sqrt{p_c-q_c} - t_c\sqrt{q_c} - \frac{y_c}{\sqrt{\alpha}}\right)^2\right]\mathrm{D}z_c\right]^n \mathrm{D}t_c\,\mathrm{d}y_c \\
&= \sqrt{\frac{\left(1+\frac{\alpha}{\sigma^2}(p_c-q_c)\right)^{1-n}}{1+\frac{\alpha}{\sigma^2}(p_c-q_c)+n\frac{\alpha}{\sigma^2}\left(\frac{\sigma_0^2}{\alpha}+p_{0c}-2m_c+q_c\right)}}
\end{aligned}
$$

with the Gaussian measure $\mathrm{D}z = \exp(-z^2/2)\,\mathrm{d}z/\sqrt{2\pi}$.

Saddle Point Integration

Partial derivatives of

$$\frac{1}{N}\sum_{c=1}^{N}\left(G(\boldsymbol{Q}[c])-\alpha\sum_{a\leq b}\tilde{Q}_{ab}[c]Q_{ab}[c]\right)$$

with respect to m_c, q_c, p_c and p_{0c} must vanish as $N \to \infty$. Solving for E_c, F_c, G_c, and G_{0c} and letting $n \to 0$ yields for all c

$$\begin{aligned}
E_c &= \frac{1}{\sigma^2+\alpha(p_c-q_c)} \\
F_c &= \frac{\sigma_0^2+\alpha\left(p_{0c}-2m_c+q_c\right)}{[\sigma^2+\alpha(p_c-q_c)]^2} \\
G_c &= F_c - E_c \\
G_{0c} &= 0.
\end{aligned}$$

In order to proceed with the calculations, we specify a prior distribution.

Binary Prior Distribution

Consider a non-uniform binary prior

$$\mathrm{p}_a(x_{ak}) = \frac{1+t_k}{2}\,\delta(x_{ak}-1) + \frac{1-t_k}{2}\,\delta(x_{ak}+1).$$

Plugging the prior distribution into the moment generating function, we find

$$M_k(\tilde{E}_k, \tilde{F}_k, \tilde{G}_k, \tilde{G}_{0k}) = $$
$$= \frac{\int \frac{1+t_k}{2}\cosh^n\left(z\sqrt{\tilde{F}_k} + \tilde{E}_k + \frac{\lambda_k}{2}\right) + \frac{1-t_k}{2}\cosh^n\left(z\sqrt{\tilde{F}_k} + \tilde{E}_k - \frac{\lambda_k}{2}\right)\mathrm{D}z}{\cosh^n\left(\frac{\lambda_k}{2}\right)\exp\left(\frac{n\tilde{F}_k - \tilde{G}_{0k} - n\tilde{G}_k}{2}\right)}$$

with $t_k = \tanh(\lambda_k/2)$.

Binary Prior Distribution (cont'd)

By saddle point integration, partial derivations of

$$\log \prod_{k=1}^{K} M_k\left(\tilde{E}_k, \tilde{F}_k, \tilde{G}_k, \tilde{G}_{0k}\right) - \alpha \sum_{c=1}^{N} nE_c m_c + \frac{n(n-1)}{2} F_c q_c + \frac{G_{0c} p_{0c}}{2} + \frac{n}{2} G_c p_c$$

with respect to E_c, F_c, G_c, G_{0c} must vanish for all c as $N \to \infty$.

An explicit calculation of these derivatives gives

$$m_c = \frac{1}{K} \sum_{k=1}^{K} w_{ck}^2 \int \tfrac{1+t_k}{2} \tanh\left(z\sqrt{\tilde{F}_k} + \tilde{E}_k + \tfrac{\lambda_k}{2}\right) + \tfrac{1-t_k}{2} \tanh\left(z\sqrt{\tilde{F}_k} + \tilde{E}_k - \tfrac{\lambda_k}{2}\right) \mathrm{D}z$$

$$q_c = \frac{1}{K} \sum_{k=1}^{K} w_{ck}^2 \int \tfrac{1+t_k}{2} \tanh^2\left(z\sqrt{\tilde{F}_k} + \tilde{E}_k + \tfrac{\lambda_k}{2}\right) + \tfrac{1-t_k}{2} \tanh^2\left(z\sqrt{\tilde{F}_k} + \tilde{E}_k - \tfrac{\lambda_k}{2}\right) \mathrm{D}z$$

$$p_c = p_{0c} = \frac{1}{K} \sum_{k=1}^{K} w_{ck}^2$$

in the limit $n \to 0$.

Binary Prior Distribution (cont'd)

Collecting our previous results to evaluate the free energy, we find

$$\frac{\mathrm{F}(\boldsymbol{x})}{K} = \lim_{n\to 0}\frac{1}{K}\frac{\partial}{\partial n}\sum_{c=1}^{N}\left[-G(m_c,q_c,p_c,p_{0c}) + \alpha n E_c m_c + \frac{\alpha n(n-1)}{2}F_c q_c + \frac{\alpha n}{2}G_c p_c\right]$$

$$-\sum_{k=1}^{K}\log M_k\left(\tilde{E}_k,\tilde{F}_k,\tilde{G}_k,0\right)$$

$$= \frac{1}{2K}\sum_{c=1}^{N}\left[\log\left(1+\frac{\alpha}{\sigma^2}(p_c-q_c)\right) + \frac{F_c}{E_c} + 2\alpha E_c m_c - \alpha F_c q_c + \alpha G_c p_c\right]$$

$$-\frac{1}{K}\sum_{k=1}^{K}\int \frac{1+t_k}{2}\log\cosh\left(z\sqrt{\tilde{F}_k}+\tilde{E}_k+\frac{\lambda_k}{2}\right)$$

$$+\frac{1-t_k}{2}\log\cosh\left(z\sqrt{\tilde{F}_k}+\tilde{E}_k-\frac{\lambda_k}{2}\right)\mathrm{D}z + \frac{1}{2}\log\left(1-t_k^2\right) - \frac{\tilde{F}_k+\tilde{G}_k}{2}.$$

The six macroscopic parameters $E_c, F_c, G_c, m_c, q_c, p_c$ are implicitly given by the saddle point equations. This system of equations can only be solved numerically.

In case of multiple solutions, the correct solution is that one which minimizes the free energy.

Matched Binary Prior Distribution

Specializing our result to the matched detector by letting $\sigma \to \sigma_0$, we have $F_c \to E_c$, $G_c \to G_{0c}$, $q_c \to m_c$. This makes the free energy simplify to

$$\begin{aligned}
\frac{\mathrm{F}(\boldsymbol{x})}{K} &= \frac{1}{2K}\sum_{c=1}^{N}\left[\log\left(1+\frac{\alpha}{\sigma_0^2}(p_{0c}-m_c)\right)+1+\alpha E_c m_c\right]-\frac{1}{K}\sum_{k=1}^{K}\log\sqrt{1-t_k^2}-\frac{\tilde{E}_k}{2}\\
&\quad+\int \tfrac{1+t_k}{2}\log\cosh\left(z\sqrt{\tilde{E}_k}+\tilde{E}_k+\tfrac{\lambda_k}{2}\right)+\tfrac{1-t_k}{2}\log\cosh\left(z\sqrt{\tilde{E}_k}+\tilde{E}_k-\tfrac{\lambda_k}{2}\right)\mathrm{D}z\\
&= \frac{1}{2K}\sum_{c=1}^{N}\left[\sigma_0^2 E_c-\log\left(\sigma_0^2 E_c\right)\right]-\frac{1}{K}\sum_{k=1}^{K}\log\sqrt{1-t_k^2}\\
&\quad+\int \tfrac{1+t_k}{2}\log\cosh\left(z\sqrt{\tilde{E}_k}+\tilde{E}_k+\tfrac{\lambda_k}{2}\right)+\tfrac{1-t_k}{2}\log\cosh\left(z\sqrt{\tilde{E}_k}+\tilde{E}_k-\tfrac{\lambda_k}{2}\right)\mathrm{D}z
\end{aligned}$$

where the macroscopic parameters E_c are given by

$$\frac{1}{E_c} = \sigma_0^2+\frac{\alpha}{K}\sum_{k=1}^{K}w_{ck}^2\left(1-t_k^2\right)\int\frac{1-\tanh\left(z\sqrt{\tilde{E}_k}+\tilde{E}_k\right)}{1-t_k^2\tanh^2\left(z\sqrt{\tilde{E}_k}+\tilde{E}_k\right)}\mathrm{D}z.$$

Equivalent AWGN Channel

Similar to the case of Gaussian priors, $\tilde{E}_k$ can be shown to be a kind of signal-to-interference and noise ratio, in the sense that the bit error probability of user k is given by

$$\Pr(\hat{x}_k \neq x_k) = \int_{\sqrt{\tilde{E}_k}}^{\infty} \mathrm{D}z.$$

An equivalent additive white Gaussian noise channel can be defined to model the multiuser interference for any prior.

MC-CDMA in Multipath Fading

Equivalent baseband vector channel in frequency domain:

$$\underset{N\times 1}{\boldsymbol{y}} = \Big(\underset{N\times K}{\boldsymbol{W}} \odot \underset{N\times K}{\boldsymbol{S}} \Big) \underset{K\times 1}{\boldsymbol{x}} + \underset{N\times 1}{\boldsymbol{n}}$$

$\boldsymbol{y}$	$\boldsymbol{W}$	$\odot$	$\boldsymbol{S}$	$\boldsymbol{x}$	$\boldsymbol{n}$
$N \times 1$	$N \times K$		$N \times K$	$K \times 1$	$N \times 1$
frequency chips	channel matrix	Hadamard product	spreading matrix	users' data	noise vector

- The noise $\boldsymbol{n}$ has i.i.d. Gaussian entries with *zero-mean* and *unit variance*.
- The columns of $\boldsymbol{S}$ are the random spreading sequences of the users.
- The columns of $\boldsymbol{W}$ are the random frequency responses of the users.

Minimum Probability of Error for MAP Detector

Maximum a-posteriori detector:

$$\hat{x}_k = \arg\max_{x_k} \Pr(x_k|\boldsymbol{y}, \boldsymbol{W})$$

In the large system limit, there is an equivalent AWGN channel with SINR $\tilde{E}_k$ such that

$$\Pr(\hat{x}_k \neq x_k|\boldsymbol{W}) = \int_{\sqrt{\tilde{E}_k}}^{\infty} Dz = \mathrm{Q}\left(\sqrt{\tilde{E}_k}\right)$$

and

$$\Pr(\hat{x}_k \neq x_k) = \mathop{\mathrm{E}}_{\boldsymbol{W}} \Pr(\hat{x}_k \neq x_k|\boldsymbol{W}) = \mathop{\mathrm{E}}_{\boldsymbol{W}} \mathrm{Q}\left(\sqrt{\tilde{E}_k}\right)$$

SINR of Equivalent AWGN Channel

For N, K large, solve the fixed-point system of equations

$$\tilde{E}_k = \frac{1}{N}\sum_{c=1}^{N} E_c w_{ck}^2$$

$$E_c = \frac{1}{\sigma_n^2 + \frac{\alpha}{K}\sum_{k=1}^{K}(1-t_k^2)w_{ck}^2 \int \frac{1-\tanh\left(z\sqrt{\tilde{E}_k}+\tilde{E}_k\right)}{1-t_k^2\tanh^2\left(z\sqrt{\tilde{E}_k}+\tilde{E}_k\right)}Dz}$$

In practice, the fading statistics obey some structure:

- Asymptotic frequency-invariance on the uplink (reverse link)
- Rank-1 statistics on the downlink (forward link)

Asymptotic Frequency Invariance (Uplink)

$$E_c = E \qquad \forall c$$

The fading is ergodic across the user population for each frequency c.

$$\tilde{E}_k = \frac{P_k}{\sigma_n^2 + \frac{\alpha}{K}\sum_{k'=1}^{K}(1-t_{k'}^2)P_{k'}\int \frac{1-\tanh\left(z\sqrt{\tilde{E}_{k'}}+\tilde{E}_{k'}\right)}{1-t_{k'}^2\tanh^2\left(z\sqrt{\tilde{E}_{k'}}+\tilde{E}_{k'}\right)}Dz}$$

with

$$P_k = \frac{1}{N}\sum_{c=1}^{N} w_{ck}^2$$

The spectrum of the received signal is white (frequency-invariant).

Full diversity is achieved.

Rank 1 Statistics (Downlink)

$$\boldsymbol{W} = \boldsymbol{f}\boldsymbol{u}^{\mathrm{T}} \qquad \Longleftrightarrow \qquad w_{ck} = f_c u_k$$

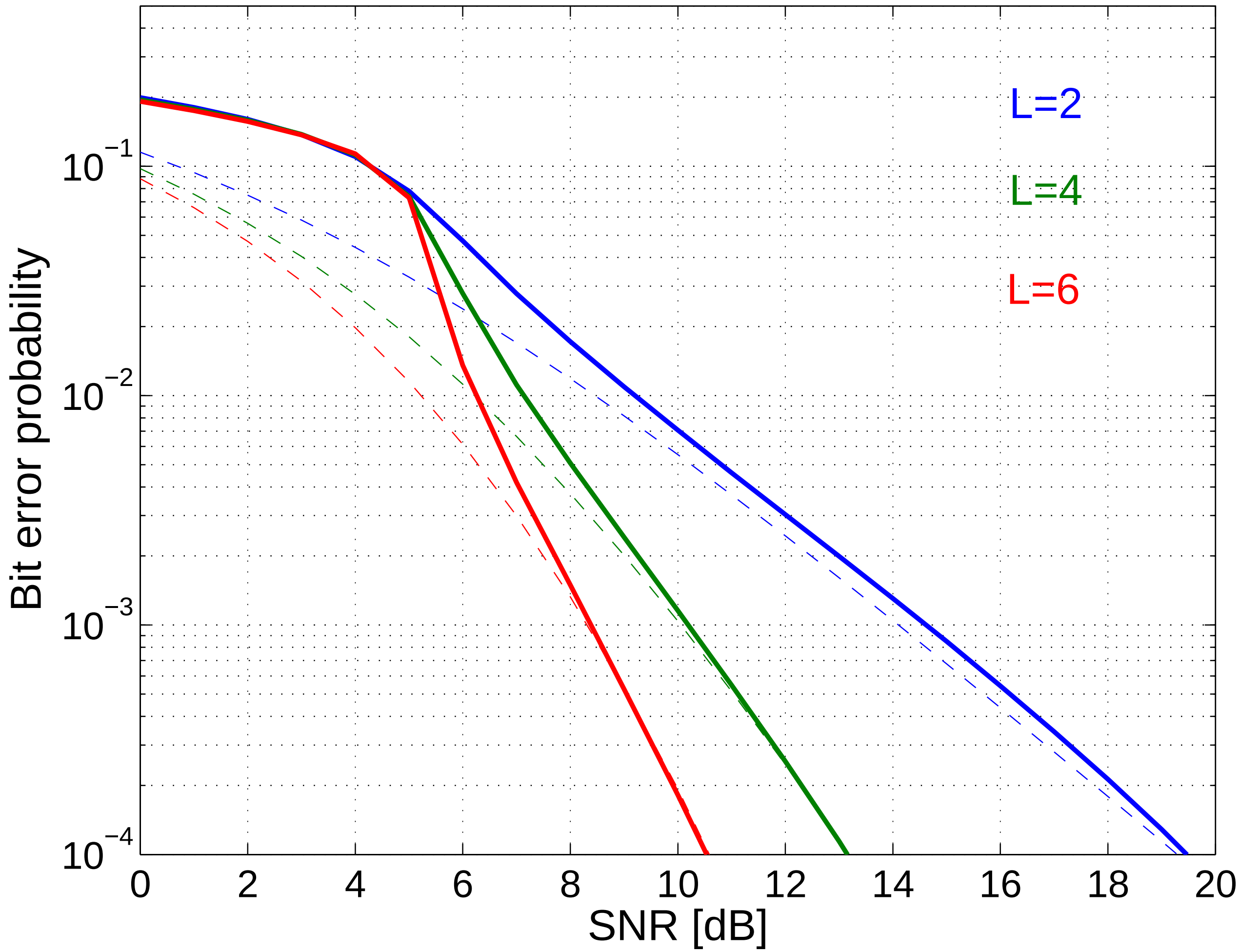
All users experience the same fading channel except for a scalar factor u_k.

$$\tilde{E}_k = \frac{u_k^2}{N} \sum_{c=1}^{N} \frac{1}{\frac{\sigma_n^2}{f_c^2} + \frac{\alpha}{K} \sum_{n=1}^{K} (1 - t_n^2) u_n^2 \int \frac{1 - \tanh\left(z\sqrt{\tilde{E}_n} + \tilde{E}_n\right)}{1 - t_n^2 \tanh^2\left(z\sqrt{\tilde{E}_n} + \tilde{E}_n\right)} Dz}$$

Full diversity is achieved.

The spectrum of the received signal is colored $\Longrightarrow$ degradation.

Uplink

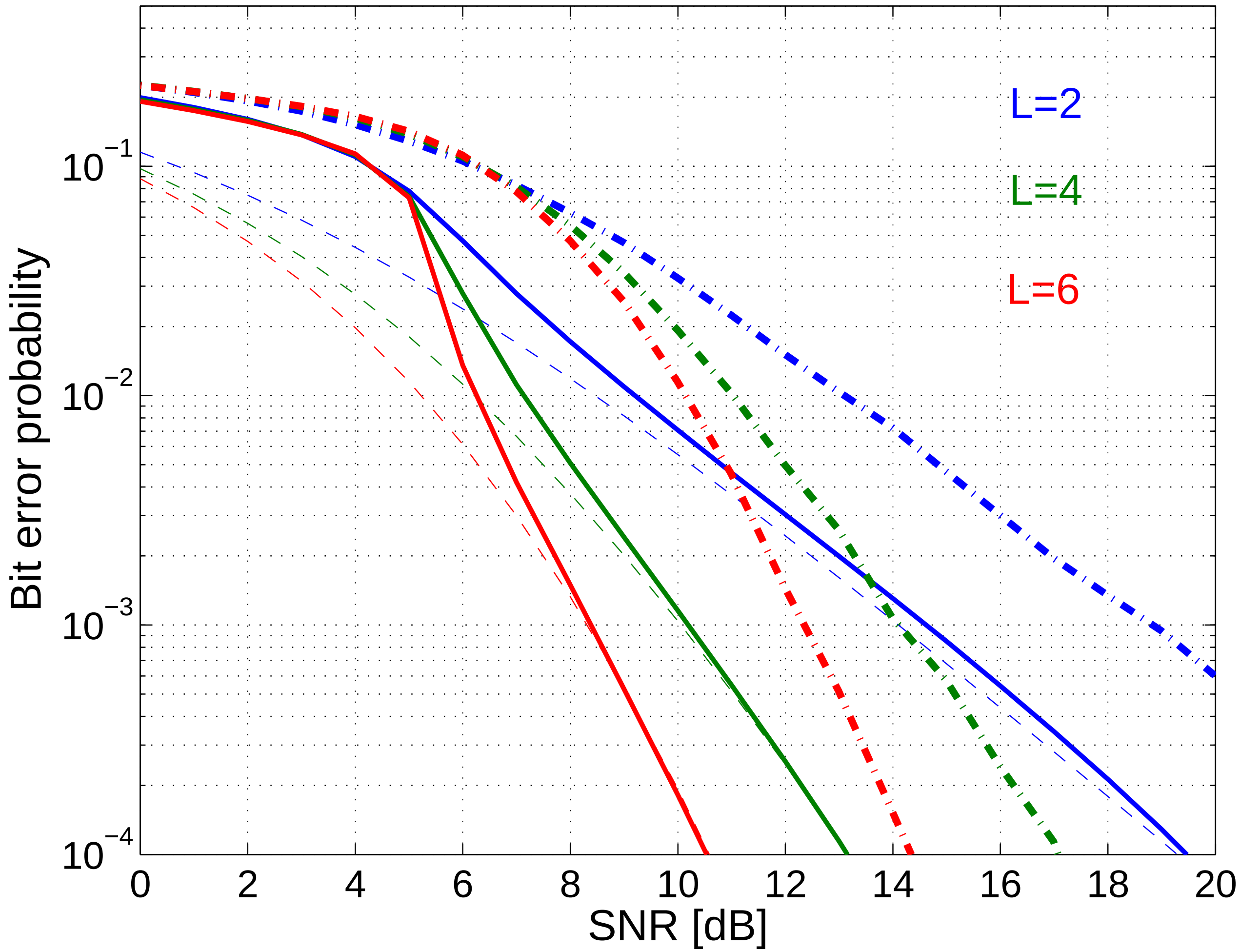


Uniform priors

L equal power paths

$$\frac{K}{N} = 1.5$$

Uplink vs. Downlink



Uniform priors

L equal power paths

$$\frac{K}{N} = 1.5$$

4 dB difference!

Vector Precoding

Let

$$E := \frac{1}{K} \min_{\boldsymbol{x} \in \mathcal{X}} \boldsymbol{x}^{\mathrm{H}} \boldsymbol{J} \boldsymbol{x}$$

with $\boldsymbol{x} \in \mathbb{C}^K$ and $\boldsymbol{J} \in \mathbb{C}^{K \times K}$.

Vector Precoding

Let

$$E := \frac{1}{K} \min_{\boldsymbol{x} \in \mathcal{X}} \boldsymbol{x}^{\mathrm{H}} \boldsymbol{J} \boldsymbol{x}$$

with $\boldsymbol{x} \in \mathbb{C}^K$ and $\boldsymbol{J} \in \mathbb{C}^{K \times K}$.

Example 1:

$$\mathcal{X} = \{\boldsymbol{x} : \boldsymbol{x}^{\mathrm{H}} \boldsymbol{x} = K\} \quad \Longrightarrow \quad E = \min \lambda(\boldsymbol{J})$$

Vector Precoding

Let

$$E := \frac{1}{K} \min_{\boldsymbol{x} \in \mathcal{X}} \boldsymbol{x}^{\mathrm{H}} \boldsymbol{J} \boldsymbol{x}$$

with $\boldsymbol{x} \in \mathbb{C}^K$ and $\boldsymbol{J} \in \mathbb{C}^{K \times K}$.

Example 1:

$$\mathcal{X} = \{\boldsymbol{x} : \boldsymbol{x}^{\mathrm{H}} \boldsymbol{x} = K\} \quad \Longrightarrow \quad E = \min \lambda(\boldsymbol{J})$$

for Wishart matrix $\longrightarrow [1 - \sqrt{\alpha}]_+^2$

Vector Precoding

Let

$$E := \frac{1}{K} \min_{\boldsymbol{x} \in \mathcal{X}} \boldsymbol{x}^{\mathrm{H}} \boldsymbol{J} \boldsymbol{x}$$

with $\boldsymbol{x} \in \mathbb{C}^K$ and $\boldsymbol{J} \in \mathbb{C}^{K \times K}$.

Example 1:

$$\mathcal{X} = \{\boldsymbol{x} : \boldsymbol{x}^{\mathrm{H}} \boldsymbol{x} = K\} \quad \Longrightarrow \quad E = \min \lambda(\boldsymbol{J})$$

for Wishart matrix $\longrightarrow [1 - \sqrt{\alpha}]_+^2$

Example 2:

$$\mathcal{X} = \{x : x^2 = 1\}^K \quad \Longrightarrow \quad ???$$

Vector Precoding

Let

$$E := \frac{1}{K} \min_{\boldsymbol{x} \in \mathcal{X}} \boldsymbol{x}^{\mathrm{H}} \boldsymbol{J} \boldsymbol{x}$$

with $\boldsymbol{x} \in \mathbb{C}^K$ and $\boldsymbol{J} \in \mathbb{C}^{K \times K}$.

Example 1:

$$\mathcal{X} = \{\boldsymbol{x} : \boldsymbol{x}^{\mathrm{H}} \boldsymbol{x} = K\} \quad \Longrightarrow \quad E = \min \lambda(\boldsymbol{J})$$

for Wishart matrix $\longrightarrow [1 - \sqrt{\alpha}]_+^2$

Example 2:

$$\mathcal{X} = \{x : x^2 = 1\}^K \quad \Longrightarrow \quad ???$$

General product set:

$$\mathcal{X} = \{x_1 \in \mathcal{B}_1\} \times \cdots \times \{x_K \in \mathcal{B}_K\} \quad \Longrightarrow \quad ???$$

Zero Temperature Formulation

Quadratic programming is the problem of finding the zero temperature limit (ground state energy) of a quadratic Hamiltonian.

The quadratic form is written as a zero temperature limit

$$E = -\lim_{\beta\to\infty} \frac{1}{\beta K} \log \sum_{\boldsymbol{x}\in\mathcal{X}} \mathrm{e}^{-\beta K \,\mathrm{Tr}(\boldsymbol{J}\boldsymbol{x}\boldsymbol{x}^{\mathrm{H}})}$$

with $\frac{1}{\beta}$ denoting temperature.

Zero Temperature Formulation

Quadratic programming is the problem of finding the zero temperature limit (ground state energy) of a quadratic Hamiltonian.

The quadratic form is written as a zero temperature limit

$$
\begin{aligned}
E &= -\lim_{\beta\to\infty} \frac{1}{\beta K} \log \sum_{\boldsymbol{x}\in\mathcal{X}} \mathrm{e}^{-\beta K \boldsymbol{x}^{\mathrm{H}} \boldsymbol{J} \boldsymbol{x}} \\
&\longrightarrow -\lim_{\beta\to\infty} \lim_{K\to\infty} \mathop{\mathrm{E}}_{\boldsymbol{J}} \frac{1}{\beta K} \log \sum_{\boldsymbol{x}\in\mathcal{X}} \mathrm{e}^{-\beta K \,\mathrm{Tr}(\boldsymbol{J}\boldsymbol{x}\boldsymbol{x}^{\mathrm{H}})}
\end{aligned}
$$

with $\frac{1}{\beta}$ denoting temperature and assumed to be self-averaging.

We have converted the optimization into the limit of an analytic function.

Replica Continuity

We want

$$\lim_{K\to\infty} \frac{1}{K} \mathop{\mathrm{E}}_{\boldsymbol{J}} \log \sum_{\boldsymbol{x}\in\mathcal{X}} \mathrm{e}^{-\beta K \boldsymbol{x}^{\mathrm{H}} \boldsymbol{J} \boldsymbol{x}} = \lim_{K\to\infty} \lim_{n\to 0} \frac{1}{nK} \log \mathop{\mathrm{E}}_{\boldsymbol{J}} \left(\sum_{\boldsymbol{x}\in\mathcal{X}} \mathrm{e}^{-\beta K \operatorname{Tr}(\boldsymbol{J}\boldsymbol{x}\boldsymbol{x}^{\mathrm{H}})} \right)^{n}$$

Replica Continuity

We want

$$\lim_{K\to\infty} \frac{1}{K} \mathop{\mathrm{E}}_{\boldsymbol{J}} \log \sum_{\boldsymbol{x}\in\mathcal{X}} \mathrm{e}^{-\beta K \boldsymbol{x}^{\mathrm{H}}\boldsymbol{J}\boldsymbol{x}} = \lim_{K\to\infty} \lim_{n\to 0} \frac{1}{nK} \log \mathop{\mathrm{E}}_{\boldsymbol{J}} \left(\sum_{\boldsymbol{x}\in\mathcal{X}} \mathrm{e}^{-\beta K \,\mathrm{Tr}(\boldsymbol{J}\boldsymbol{x}\boldsymbol{x}^{\mathrm{H}})} \right)^{n}$$

$$= \lim_{K\to\infty} \lim_{n\to 0} \frac{1}{nK} \log \mathop{\mathrm{E}}_{\boldsymbol{J}} \prod_{a=1}^{n} \sum_{\boldsymbol{x}_a\in\mathcal{X}} \mathrm{e}^{-\beta K \,\mathrm{Tr}\left(\boldsymbol{J}\boldsymbol{x}_a\boldsymbol{x}_a^{\mathrm{H}}\right)}$$

Replica Continuity

We want

$$\lim_{K\to\infty} \frac{1}{K} \mathop{\mathrm{E}}_{\boldsymbol{J}} \log \sum_{\boldsymbol{x}\in\mathcal{X}} \mathrm{e}^{-\beta K \boldsymbol{x}^{\mathrm{H}} \boldsymbol{J} \boldsymbol{x}} = \lim_{K\to\infty} \lim_{n\to 0} \frac{1}{nK} \log \mathop{\mathrm{E}}_{\boldsymbol{J}} \left(\sum_{\boldsymbol{x}\in\mathcal{X}} \mathrm{e}^{-\beta K \,\mathrm{Tr}(\boldsymbol{J}\boldsymbol{x}\boldsymbol{x}^{\mathrm{H}})} \right)^n$$

$$= \lim_{K\to\infty} \lim_{n\to 0} \frac{1}{nK} \log \mathop{\mathrm{E}}_{\boldsymbol{J}} \prod_{a=1}^{n} \sum_{\boldsymbol{x}_a\in\mathcal{X}} \mathrm{e}^{-\beta K \,\mathrm{Tr}\left(\boldsymbol{J}\boldsymbol{x}_a\boldsymbol{x}_a^{\mathrm{H}}\right)}$$

$$= \lim_{K\to\infty} \lim_{n\to 0} \frac{1}{nK} \log \mathop{\mathrm{E}}_{\boldsymbol{J}} \sum_{\boldsymbol{x}_1\in\mathcal{X}} \cdots \sum_{\boldsymbol{x}_n\in\mathcal{X}} \mathrm{e}^{-K \,\mathrm{Tr}\left(\boldsymbol{J}\beta \sum\limits_{a=1}^{n} \boldsymbol{x}_a\boldsymbol{x}_a^{\mathrm{H}}\right)}$$

Replica Continuity

We want

$$\lim_{K\to\infty}\frac{1}{K}\mathop{\mathrm{E}}_{\boldsymbol{J}}\log\sum_{\boldsymbol{x}\in\mathcal{X}} \mathrm{e}^{-\beta K \boldsymbol{x}^{\mathrm{H}}\boldsymbol{J}\boldsymbol{x}} = \lim_{K\to\infty}\lim_{n\to 0}\frac{1}{nK}\log\mathop{\mathrm{E}}_{\boldsymbol{J}}\left(\sum_{\boldsymbol{x}\in\mathcal{X}} \mathrm{e}^{-\beta K\,\mathrm{Tr}(\boldsymbol{J}\boldsymbol{x}\boldsymbol{x}^{\mathrm{H}})}\right)^n$$

$$= \lim_{K\to\infty}\lim_{n\to 0}\frac{1}{nK}\log\mathop{\mathrm{E}}_{\boldsymbol{J}}\prod_{a=1}^{n}\sum_{\boldsymbol{x}_a\in\mathcal{X}} \mathrm{e}^{-\beta K\,\mathrm{Tr}\left(\boldsymbol{J}\boldsymbol{x}_a\boldsymbol{x}_a^{\mathrm{H}}\right)}$$

$$= \lim_{K\to\infty}\lim_{n\to 0}\frac{1}{nK}\log\mathop{\mathrm{E}}_{\boldsymbol{J}}\sum_{\boldsymbol{x}_1\in\mathcal{X}}\cdots\sum_{\boldsymbol{x}_n\in\mathcal{X}} \mathrm{e}^{-K\,\mathrm{Tr}\left(\boldsymbol{J}\beta\sum_{a=1}^{n}\boldsymbol{x}_a\boldsymbol{x}_a^{\mathrm{H}}\right)}$$

$$= \lim_{K\to\infty}\lim_{n\to 0}\frac{1}{nK}\log\mathop{\mathrm{E}}_{\boldsymbol{J}}\sum_{\boldsymbol{X}\in\mathcal{X}^n} \mathrm{e}^{-K\,\mathrm{Tr}\left(\boldsymbol{J}\beta\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}\right)}$$

Replica Continuity

We want

$$\lim_{K\to\infty}\frac{1}{K}\mathop{\mathrm{E}}_{\boldsymbol{J}}\log\sum_{\boldsymbol{x}\in\mathcal{X}}\mathrm{e}^{-\beta K\boldsymbol{x}^{\mathrm{H}}\boldsymbol{J}\boldsymbol{x}} = \lim_{K\to\infty}\lim_{n\to 0}\frac{1}{nK}\log\mathop{\mathrm{E}}_{\boldsymbol{J}}\left(\sum_{\boldsymbol{x}\in\mathcal{X}}\mathrm{e}^{-\beta K\,\mathrm{Tr}(\boldsymbol{J}\boldsymbol{x}\boldsymbol{x}^{\mathrm{H}})}\right)^{n}$$

$$= \lim_{K\to\infty}\lim_{n\to 0}\frac{1}{nK}\log\mathop{\mathrm{E}}_{\boldsymbol{J}}\prod_{a=1}^{n}\sum_{\boldsymbol{x}_a\in\mathcal{X}}\mathrm{e}^{-\beta K\,\mathrm{Tr}\left(\boldsymbol{J}\boldsymbol{x}_a\boldsymbol{x}_a^{\mathrm{H}}\right)}$$

$$= \lim_{K\to\infty}\lim_{n\to 0}\frac{1}{nK}\log\mathop{\mathrm{E}}_{\boldsymbol{J}}\sum_{\boldsymbol{x}_1\in\mathcal{X}}\cdots\sum_{\boldsymbol{x}_n\in\mathcal{X}}\mathrm{e}^{-K\,\mathrm{Tr}\left(\boldsymbol{J}\beta\sum_{a=1}^{n}\boldsymbol{x}_a\boldsymbol{x}_a^{\mathrm{H}}\right)}$$

$$= \lim_{K\to\infty}\lim_{n\to 0}\frac{1}{nK}\log\mathop{\mathrm{E}}_{\boldsymbol{J}}\sum_{\boldsymbol{X}\in\mathcal{X}^n}\mathrm{e}^{-K\,\mathrm{Tr}\left(\boldsymbol{J}\beta\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}\right)}$$

$$= \lim_{K\to\infty}\lim_{n\to 0}\frac{1}{nK}\log\sum_{\boldsymbol{X}\in\mathcal{X}^n}\exp\left[-K\sum_{a=1}^{n}\int_0^{\lambda_a(\beta\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}/K)}R_{\boldsymbol{J}}(-w)\,\mathrm{d}w\right]$$

Just a Change of Variables

Define a matrix of so-called *quenched* random variables

$$\boldsymbol{Q} := \frac{\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}}{K}$$

and find

$$\lim_{K\to\infty} E = \lim_{\beta\to\infty}\lim_{n\to 0}\lim_{K\to\infty}\frac{1}{n\beta K}\log\sum_{\boldsymbol{X}\in\mathcal{X}^n} \mathrm{e}^{K\sum\limits_{a=1}^{n}\int\limits_0^{\beta\lambda_a(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}/K)} R(w)\mathrm{d}w}$$

$$= \lim_{\beta\to\infty}\lim_{n\to 0}\lim_{K\to\infty}\frac{1}{n\beta K}\log\int\limits_{\mathbb{R}^{n\times n}} \underbrace{\mathrm{e}^{K\sum\limits_{a=1}^{n}\int\limits_0^{\beta\lambda_a(\boldsymbol{Q})} R(w)\mathrm{d}w}}_{=:\mathrm{e}^{KG(\boldsymbol{Q})}} \underbrace{\sum_{\boldsymbol{X}\in\mathcal{X}^n}\delta\left(K\boldsymbol{Q}-\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}\right)}_{=:\mathrm{e}^{KI(\boldsymbol{Q})}}\mathrm{d}\boldsymbol{Q}$$

Next, we consider $G(\boldsymbol{Q})$ and $I(\boldsymbol{Q})$ separately.

Exponentials are Best

Write the Dirac measure as its inverse Laplace transform

$$\delta(Q) = \oint_{\mathcal{C}} \mathrm{e}^{Q\tilde{Q}} \, \frac{\mathrm{d}\tilde{Q}}{2\pi \mathrm{j}}$$

for some appropriate contour $\mathcal{C}$.

Exponentials are Best

Write the Dirac measure as its inverse Laplace transform

$$\delta(Q) = \oint\limits_{\mathcal{C}} \mathrm{e}^{Q\tilde{Q}} \, \frac{\mathrm{d}\tilde{Q}}{2\pi \mathrm{j}}$$

for some appropriate contour $\mathcal{C}$.

In $n \times n$ dimensions, we have

$$\delta(\boldsymbol{Q}) = \oint\limits_{\mathcal{C}^{n\times n}} \mathrm{e}^{\operatorname{tr}(\boldsymbol{Q}\tilde{\boldsymbol{Q}})} \, \frac{\mathrm{d}\tilde{\boldsymbol{Q}}}{(2\pi \mathrm{j})^{n^2}}.$$

Why Exponentials are Best

$$\begin{aligned} e^{KI(\boldsymbol{Q})} &= \sum_{\boldsymbol{X}\in\mathcal{X}^n} \delta\left(K\boldsymbol{Q} - \boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}\right) \\ &= \sum_{\boldsymbol{X}\in\mathcal{X}^n} \int\limits_{\mathcal{C}^{n\times n}} e^{K\,\mathrm{tr}\left(\boldsymbol{Q}\tilde{\boldsymbol{Q}}\right)} e^{-\mathrm{tr}\left(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}\tilde{\boldsymbol{Q}}\right)} \frac{\mathrm{d}\tilde{\boldsymbol{Q}}}{(2\pi \mathrm{j})^{n^2}} \end{aligned}$$

Why Exponentials are Best

$$
\begin{aligned}
\mathrm{e}^{KI(\boldsymbol{Q})} &= \sum_{\boldsymbol{X}\in\mathcal{X}^n} \delta\left(K\boldsymbol{Q} - \boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}\right) \\
&= \sum_{\boldsymbol{X}\in\mathcal{X}^n} \int_{\mathcal{C}^{n\times n}} \mathrm{e}^{K\,\mathrm{tr}\left(\boldsymbol{Q}\tilde{\boldsymbol{Q}}\right)} \mathrm{e}^{-\mathrm{tr}\left(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}\tilde{\boldsymbol{Q}}\right)} \frac{\mathrm{d}\tilde{\boldsymbol{Q}}}{(2\pi\mathrm{j})^{n^2}} \\
&= \int_{\mathcal{C}^{n\times n}} \mathrm{e}^{K\,\mathrm{tr}\left(\boldsymbol{Q}\tilde{\boldsymbol{Q}}\right)} \sum_{\boldsymbol{X}\in\mathcal{X}^n} \mathrm{e}^{-\mathrm{tr}\left(\boldsymbol{X}\tilde{\boldsymbol{Q}}\boldsymbol{X}^{\mathrm{H}}\right)} \frac{\mathrm{d}\tilde{\boldsymbol{Q}}}{(2\pi\mathrm{j})^{n^2}} \\
&= \int_{\mathcal{C}^{n\times n}} \mathrm{e}^{K\,\mathrm{tr}\left(\boldsymbol{Q}\tilde{\boldsymbol{Q}}\right)} \sum_{\mathbf{x}_1\in\mathcal{B}_1^n} \cdots \sum_{\mathbf{x}_K\in\mathcal{B}_K^n} \mathrm{e}^{-\sum\limits_{k=1}^{K} \mathrm{tr}\left(\mathbf{x}_k\tilde{\boldsymbol{Q}}\,\mathbf{x}_k^{\mathrm{H}}\right)} \frac{\mathrm{d}\tilde{\boldsymbol{Q}}}{(2\pi\mathrm{j})^{n^2}}
\end{aligned}
$$

Why Exponentials are Best

$$
\begin{aligned}
\mathrm{e}^{KI(\boldsymbol{Q})} &= \sum_{\boldsymbol{X}\in\mathcal{X}^n} \delta\left(K\boldsymbol{Q} - \boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}\right) \\
&= \sum_{\boldsymbol{X}\in\mathcal{X}^n} \int_{\mathcal{C}^{n\times n}} \mathrm{e}^{K\,\mathrm{tr}\left(\boldsymbol{Q}\tilde{\boldsymbol{Q}}\right)} \mathrm{e}^{-\mathrm{tr}\left(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}\tilde{\boldsymbol{Q}}\right)} \frac{\mathrm{d}\tilde{\boldsymbol{Q}}}{(2\pi\mathrm{j})^{n^2}} \\
&= \int_{\mathcal{C}^{n\times n}} \mathrm{e}^{K\,\mathrm{tr}\left(\boldsymbol{Q}\tilde{\boldsymbol{Q}}\right)} \sum_{\boldsymbol{X}\in\mathcal{X}^n} \mathrm{e}^{-\mathrm{tr}\left(\boldsymbol{X}\tilde{\boldsymbol{Q}}\boldsymbol{X}^{\mathrm{H}}\right)} \frac{\mathrm{d}\tilde{\boldsymbol{Q}}}{(2\pi\mathrm{j})^{n^2}} \\
&= \int_{\mathcal{C}^{n\times n}} \mathrm{e}^{K\,\mathrm{tr}\left(\boldsymbol{Q}\tilde{\boldsymbol{Q}}\right)} \sum_{\mathbf{x}_1\in\mathcal{B}_1^n} \cdots \sum_{\mathbf{x}_K\in\mathcal{B}_K^n} \mathrm{e}^{-\sum\limits_{k=1}^{K}\mathrm{tr}\left(\mathbf{x}_k\tilde{\boldsymbol{Q}}\,\mathbf{x}_k^{\mathrm{H}}\right)} \frac{\mathrm{d}\tilde{\boldsymbol{Q}}}{(2\pi\mathrm{j})^{n^2}} \\
&= \int_{\mathcal{C}^{n\times n}} \mathrm{e}^{K\,\mathrm{tr}\left(\boldsymbol{Q}\tilde{\boldsymbol{Q}}\right)} \prod_{k=1}^{K} \underbrace{\sum_{\mathbf{x}\in\mathcal{B}_k^n} \mathrm{e}^{-K\,\mathrm{tr}\left(\mathbf{x}^{\mathrm{H}}\mathbf{x}\tilde{\boldsymbol{Q}}\right)}}_{=:M_k(\tilde{\boldsymbol{Q}})} \frac{\mathrm{d}\tilde{\boldsymbol{Q}}}{(2\pi\mathrm{j})^{n^2}}
\end{aligned}
$$

Now, the problem is no longer exponential in K.

Replica Symmetry

The integrals over $\boldsymbol{Q}$ and $\tilde{\boldsymbol{Q}}$ shall be solved by saddle point integration, but finding for the optimal matrices $\boldsymbol{Q}$ and $\tilde{\boldsymbol{Q}}$ seems a hopeless task.

So we make an ansatz and hope that it will work.

$$\boldsymbol{Q} \triangleq \begin{bmatrix} q+\frac{\chi}{\beta} & q & \cdots & q & q \\ q & q+\frac{\chi}{\beta} & \ddots & q & q \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ q & q & \ddots & q+\frac{\chi}{\beta} & q \\ q & q & \cdots & q & q+\frac{\chi}{\beta} \end{bmatrix}$$

with some macroscopic parameters q and χ.

We distinguish the cross-correlation between different replicas and the autocorrelation of an individual replica.

Twice Holds Better

We apply the same idea to the correlation variables in the Laplace domain and set (with a modest amount of foresight)

$$\begin{aligned} \tilde{Q}_{ab} &= \beta^2 f^2 \qquad \forall a \neq b \\ \tilde{Q}_{aa} &= \beta^2 f^2 - \beta e \qquad \forall a. \end{aligned}$$

For the evaluation of $G(\boldsymbol{Q})$, we can use replica symmetry to explicitly calculate the eigenvalues λ_i. Considerations of linear algebra lead to the conclusion that the eigenvalues χ and $\chi + \beta n q$ occur with multiplicities $n-1$ and 1, respectively. We get

$$G(q,\chi) = (n-1)\int\limits_0^{\chi} R(w)\,\mathrm{d}w + \int\limits_0^{\chi+\beta n q} R(w)\,\mathrm{d}w.$$

Saddle Point Integration

Due to saddle point integration, the derivatives of

$$G(q,\chi) + \mathrm{tr}(\tilde{\boldsymbol{Q}}\boldsymbol{Q})$$

with respect to q and χ must vanish as $K \to \infty$. The assumption of replica symmetry leads to

$$\mathrm{tr}(\tilde{\boldsymbol{Q}}\boldsymbol{Q}) = n(n-1)\beta^2 f^2 q + n\left(\beta f^2 - e\right)(\beta q + \chi)\,.$$

Taking derivatives yields

$$\begin{aligned} n\,R(\chi+\beta n q) + n(n-1)\beta f^2 + n\left(\beta f^2 - e\right) &= 0 \\ (n-1)\,R(\chi) + R(\chi + 2\beta n q) + n\left(\beta f^2 - e\right) &= 0 \end{aligned}$$

and solving for e and f gives

$$\begin{aligned} e &= R(\chi) \\ f &= \sqrt{\frac{R(\chi) - R(\chi+\beta n q)}{\beta n}} \quad \xrightarrow{n\to 0} \quad \sqrt{qR'(\chi)}. \end{aligned}$$

The Taming of the Shrew

With replica symmetry, one function still resists the limit $n \to 0$

$$M_k(e, f) = \sum_{\mathbf{x} \in \mathcal{B}_k^n} \mathrm{e}^{\beta^2 f^2 \left| \sum_{a=1}^{n} x_a \right|^2 - \sum_{a=1}^{n} \beta e |x_a|^2}$$

The Taming of the Shrew

With replica symmetry, one function still resists the limit $n \to 0$

$$M_k(e,f) = \sum_{\mathbf{x}\in\mathcal{B}_k^n} \mathrm{e}^{\beta^2 f^2 \left|\sum\limits_{a=1}^n x_a\right|^2 - \sum\limits_{a=1}^n \beta e |x_a|^2}$$

The *Hubbard-Stratonovich transform*

$$\mathrm{e}^{|x|^2} = \int_{\mathbb{C}} \mathrm{e}^{2\Re\{x^* z\}} \underbrace{\mathrm{e}^{-|z|^2}\frac{\mathrm{d}z}{\pi}}_{=:\mathrm{D}z}$$

tames it to read

$$M_k(e,f) = \sum_{\mathbf{x}\in\mathcal{B}_k^n} \int \mathrm{e}^{\beta\sum\limits_{a=1}^n 2f\Re\{x_a^* z\} - e|x_a|^2} \mathrm{D}z = \int \left(\sum_{x\in\mathcal{B}_k} \mathrm{e}^{2\beta f\Re\{x^* z\} - \beta e|x|^2} \right)^n \mathrm{D}z.$$

The Law of Large Numbers

For $K \to \infty$, we have by the law of large numbers

$$\log \prod_{k=1}^{K} M_k(e,f) = \frac{1}{K} \sum_{k=1}^{K} \log M_k(e,f)$$

$$\to \int \log \int \left(\sum_{x \in \mathcal{B}} \mathrm{e}^{2\beta f \Re\{z^* x\} - \beta e |x|^2} \right)^n \mathrm{D}z \mathrm{dP}(\mathcal{B})$$

$$=: \log M(e,f)$$

Saddle Point Integration Twice More

Partial derivatives of

$$\log M(e,f) - \operatorname{tr}(\tilde{\boldsymbol{Q}}\boldsymbol{Q})$$

with respect to f and e must vanish as $K \to \infty$. Thus,

$$\chi = \frac{1}{\sqrt{qR'(-\chi)}} \iint \frac{\sum_{x\in\mathcal{B}} \Re\{z^* x\} \mathrm{e}^{\beta 2\sqrt{qR'(-\chi)}\Re\{z^*x\} - \beta R(-\chi)|x|^2}}{\sum_{x\in\mathcal{B}} \mathrm{e}^{\beta 2\sqrt{qR'(-\chi)}\Re\{z^*x\} - \beta R(-\chi)|x|^2}} \mathrm{D}z\,\mathrm{dP}(\mathcal{B})$$

$$q = \iint \frac{\sum_{x\in\mathcal{B}} |x|^2 \mathrm{e}^{\beta 2\sqrt{qR'(-\chi)}\Re\{z^*x\} - \beta R(-\chi)|x|^2}}{\sum_{x\in\mathcal{B}} \mathrm{e}^{\beta 2\sqrt{qR'(-\chi)}\Re\{z^*x\} - \beta R(-\chi)|x|^2}} \mathrm{D}z\,\mathrm{dP}(\mathcal{B}) - \frac{\chi}{\beta}.$$

For $\beta \to \infty$, saddle point integration yields

$$\chi = \frac{1}{\sqrt{qR'(-\chi)}} \iint \Re \operatorname*{argmin}_{x\in\mathcal{B}} \left| z - \frac{R(-\chi)\,x}{\sqrt{qR'(-\chi)}} \right| z^* \mathrm{D}z\,\mathrm{dP}(\mathcal{B})$$

$$q = \iint \left| \operatorname*{argmin}_{x\in\mathcal{B}} \left| z - \frac{R(-\chi)\,x}{\sqrt{qR'(-\chi)}} \right| \right|^2 \mathrm{D}z\,\mathrm{dP}(\mathcal{B})$$

Finally

Collecting previous results, we find with replica continuity that

$$E = \lim_{\beta\to\infty} \frac{1}{\beta} \lim_{n\to 0} \frac{\partial}{\partial n} \left[(n-1) \int\limits_0^{\chi} R(-w)\,\mathrm{d}w + \int\limits_0^{\chi+\beta n q} R(-w)\,\mathrm{d}w - \log M(e,f) \right.$$
$$\left. + n(n-1) f^2\beta^2 q + n(f^2\beta - e)(\chi + \beta q) \right] \qquad (1)$$

$$= \lim_{\beta\to\infty} \frac{1}{\beta} \int\limits_0^{\chi} R(-w)\,\mathrm{d}w - \frac{\chi}{\beta} R(-\chi) + q\chi R'(-\chi)$$
$$- \frac{1}{\beta} \iint \log \sum_{x\in\mathcal{B}} \mathrm{e}^{\beta 2 f \Re\{z^* x\} - \beta e |x|^2} \mathrm{D}z\, \mathrm{dP}(\mathcal{B}). \qquad (2)$$

We use l'Hospital's rule, re-substitute χ and q, assume $0 < \chi < \infty$ and finally obtain

$$E = q\left[R(-\chi) - \chi R'(-\chi)\right].$$

1-Step Replica Symmetry Breaking (1RSB)

$$\boldsymbol{Q} \triangleq \overbrace{\phantom{q+p+\frac{\chi}{\beta} \quad q+p \quad q}}^{\frac{\mu}{\beta}\text{ columns}}$$

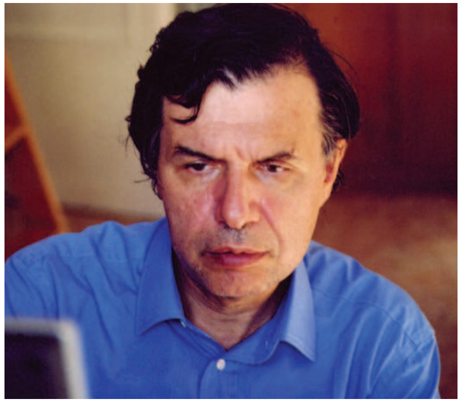
$$\boldsymbol{Q} \triangleq \begin{bmatrix} q+p+\frac{\chi}{\beta} & q+p & q & q & \cdots & q & q \\ q+p & q+p+\frac{\chi}{\beta} & q & q & \cdots & q & q \\ q & q & q+p+\frac{\chi}{\beta} & q+p & \ddots & q & q \\ q & q & q+p & q+p+\frac{\chi}{\beta} & & \vdots & \vdots \\ \vdots & \vdots & \ddots & & \ddots & q & q \\ q & q & q & \cdots & q & q+p+\frac{\chi}{\beta} & q+p \\ q & q & q & \cdots & q & q+p & q+p+\frac{\chi}{\beta} \end{bmatrix}$$

with the macroscopic parameters q, p and χ and the blocksize $\frac{\mu}{\beta}$.

1RSB Calculations

Redo, the same procedure as for RS, but now with more macroscopic parameters. The parameter μ is chosen as to extremize the free energy.

Replica symmetry breaking was introduced and solved for the semicircle law by Parisi in 1980.



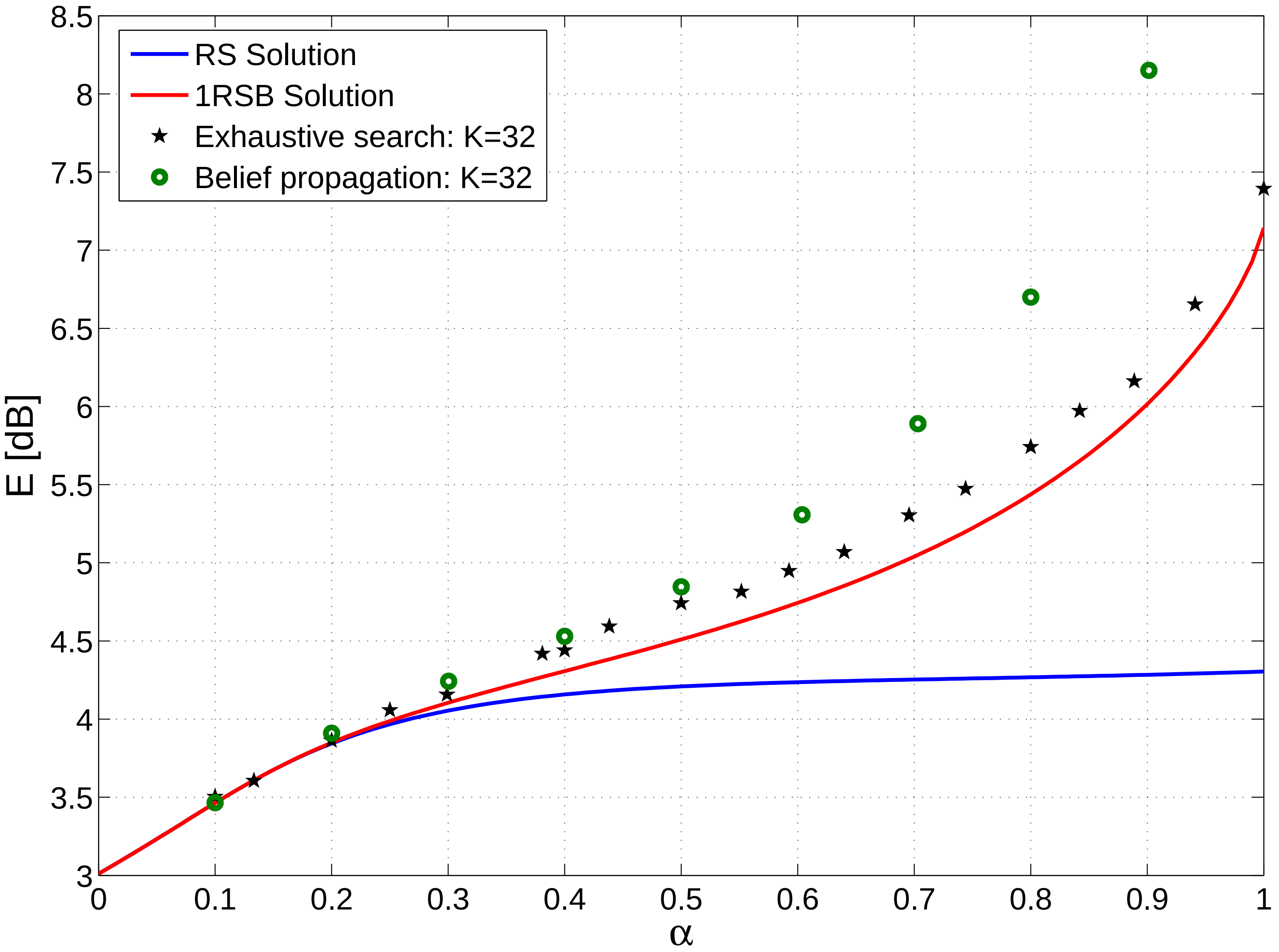
Giorgio Parisi
born in Rome in 1948

1-Step Replica Symmetry Breaking

$$E = \frac{1}{K} \min_{\boldsymbol{x} \in \mathcal{X}} \boldsymbol{x}^{\mathrm{H}} \boldsymbol{J} \boldsymbol{x}$$
$$\rightarrow \left(q + p + \frac{\chi}{\mu} \right) R_{\boldsymbol{J}}(-\chi - \mu p) - \frac{\chi}{\mu} R_{\boldsymbol{J}}(-\chi) - q(\mu p + \chi) R'_{\boldsymbol{J}}(-\chi - \mu p)$$

The macroscopic parameters q, p, χ and μ are given by 4 coupled non-linear equations (omited here).

Solving those equations numerically is a tedious and tricky task.


RS Solution
1RSB Solution
Exhaustive search: K=32
Belief propagation: K=32
E [dB]
α
8.5
8
7.5
7
6.5
6
5.5
5
4.5
4
3.5
3
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

The Meaning of Replica Symmetry Breaking

Replica symmetry means that all vectors close to the optimum have the same inner products, i.e. they differ only in few components.

- If there are multiple local extrema, many of those are quite close to each other.
- The problem can often be well approximated by iterative algorithms like belief propagation.

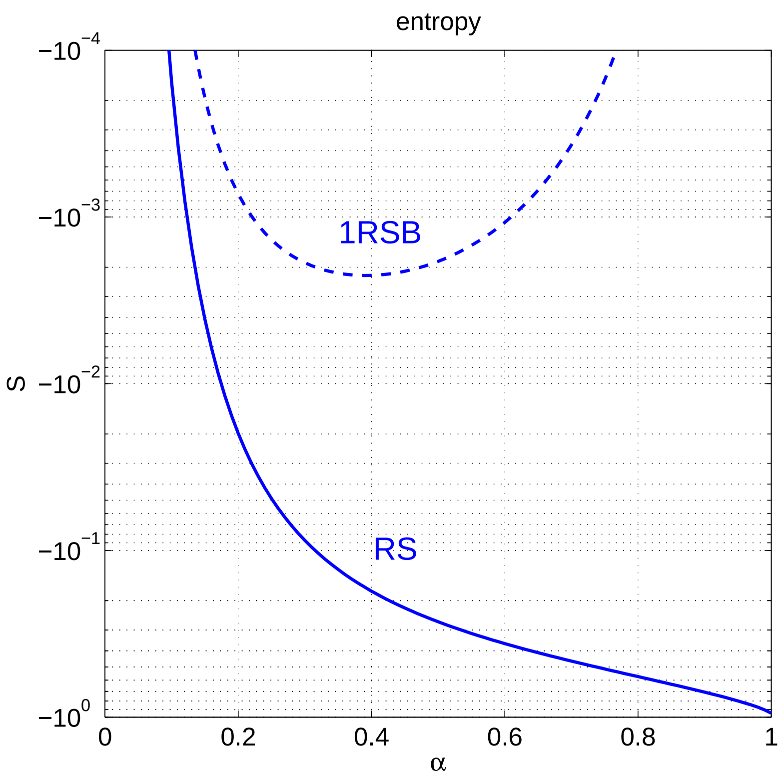
Replica symmetry breaking means that even vectors arbitrarily close to the optimum, may differ in a large portion of its components.

- There are local extrema at very different positions.
- Belief propagation is often significantly suboptimum.

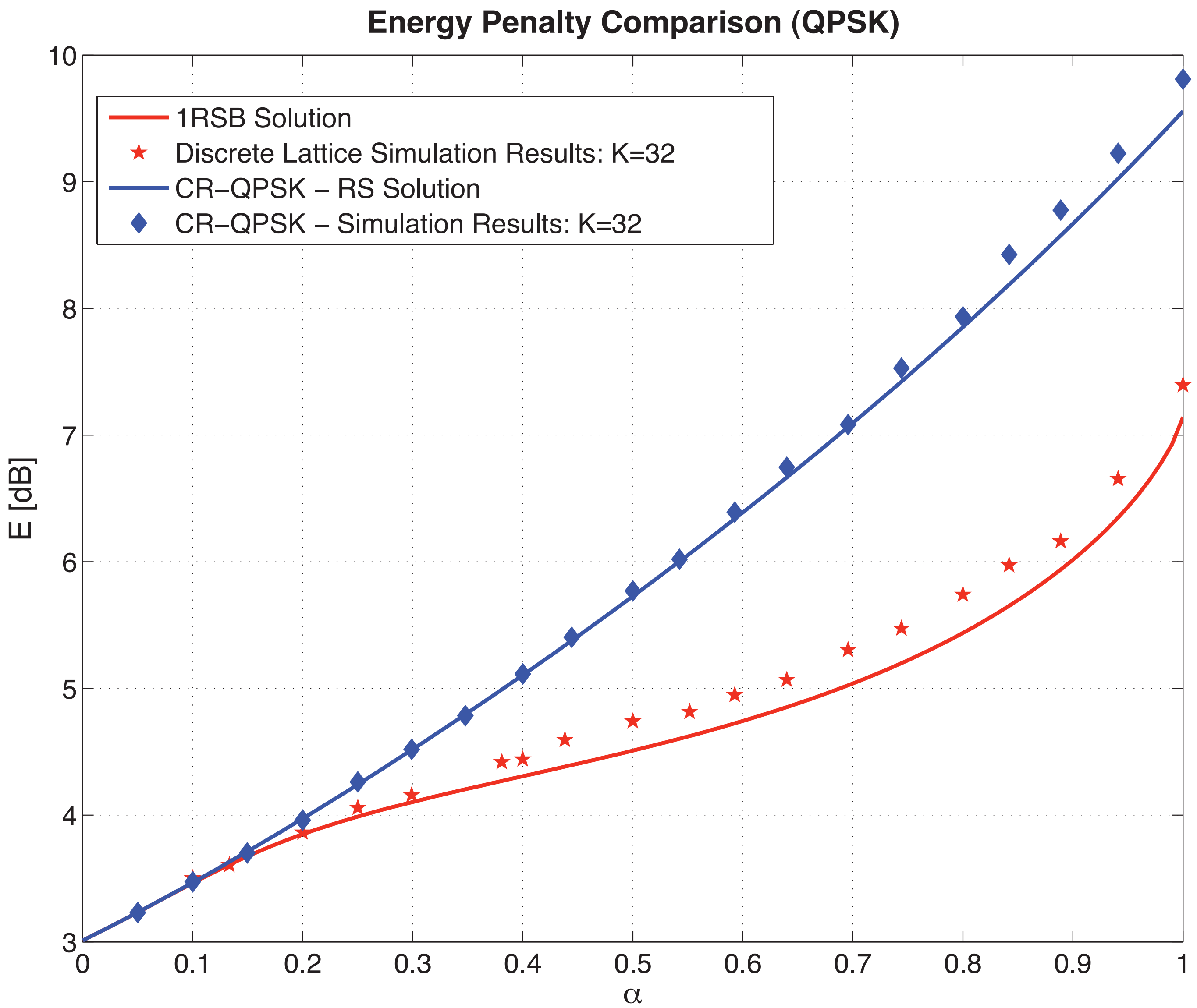
RS (breaking) ranks the difficulty of approximating an NP-hard problem, in practice.

Negative Entropy

$$S = \chi R(-\chi) - \int_0^{\chi} R(-w)\mathrm{d}w$$



The closer the entropy is to zero, the better the RSB approximation.


Energy Penalty Comparison (QPSK)
1RSB Solution
Discrete Lattice Simulation Results: K=32
CR–QPSK – RS Solution
CR–QPSK – Simulation Results: K=32
E [dB]
α
10
9
8
7
6
5
4
3
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Higher Step RSB Calculations

2RSB:

Recursively split the diagonal blocks of size $\frac{\mu}{\beta} \times \frac{\mu}{\beta}$ into subblocks of size $\frac{\mu_2}{\beta} \times \frac{\mu_2}{\beta}$ and off-diagonal blocks. Generalize p into the pair (p_1, p_2).

General RSB:

Recursively, continue this procedure until infinite order. For infinite order you get the exact result. Note that at infinite order you have to solve an infinite number of couple fixed-point equations. Sometimes, they can be written as a functional equation.