Random Matrix Theory for Wireless Communications

Prof. Dr.-Ing. Ralf R. Müller

mueller@lnt.de

April 17, 2013

Topics

Topic I: The Framework of Multiuser Detection

Multiuser detection as a canonical application in wireless communications

Topic II: The Two Theories

From classical random matrix theory to free probability theory

Topic III: The Return of Physics

Random matrix theory as a special case of statistical mechanics

Literature

[1] Sergio Verdú. *Multiuser Detection*. Cambridge University Press, New York, 1998.

[2] Antonia M. Tulino and Sergio Verdú. Random matrix theory and wireless communications. *Foundations and Trends in Communications and Information Theory*, 1(1), June 2004.

[3] Fumio Hiai and Dénes Petz. *The Semicircle Law, Free Random Variables and Entropy*. American Mathematical Society, Providence, RI, 2000.

[4] Laura Cottatellucci. *Low Complexity Multiuser Detection for Randomly Spread CDMA Systems*. PhD thesis, Technische Universität Wien, Vienna, Austria, 2005.

[5] Ralf R. Müller. Applications of large random matrices in communications engineering. In *Proc. of International Conference on Advances in the Internet, Processing, Systems, and Interdisciplinary Research (IPSI)*, Sveti Stefan, Montenegro, October 2003.

[6] Hidetoshi Nishimori. *Statistical Physics of Spin Glasses and Information Processing*. Oxford University Press, Oxford, U.K., 2001.

[7] Victor Dotsenko. *Introduction to the Replica Theory of Disordered Statistical Systems*. Cambridge University Press, Cambridge, U.K., 2001.

[8] Uffe Haagerup and Flemming Larsen. Brown's spectral distribution measure for R-diagonal elements in finite von Neumann algebras. *Journal of Functional Analysis*, 176:331–367, 2000.

[9] Ralf R. Müller. Multiuser receivers for randomly spread signals: Fundamental limits with and without decision–feedback. *IEEE Transactions on Information Theory*, 47(1):268–283, January 2001.

[10] Laura Cottatellucci, Ralf R. Müller, and Mérouane Debbah. Asynchronous CDMA systems with random spreading—Part I: Fundamental limits. *IEEE Transactions on Information Theory*, 56(4):1477–1497, April 2010.

[11] G. Parisi. A sequence of approximated solutions to the S-K model for spin glasses. *Journal of Physics A: Mathematical and General*, 13:L115–L121, 1980.

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

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

for some symbol waveforms $g_{k,\mu}(t)$ with T_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

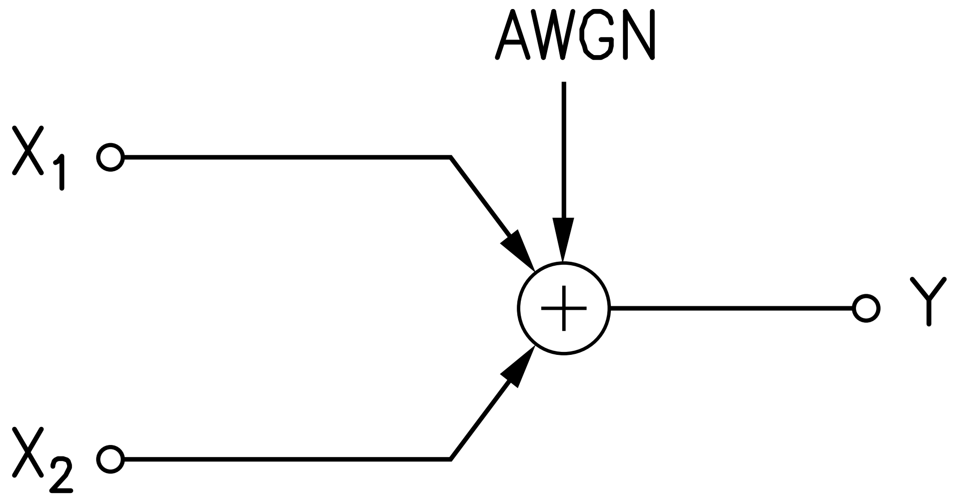
$$\psi_k(t) = \psi(t) \quad \forall\, k$$

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

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

Gaussian Multiple–Access Channel

Two users:

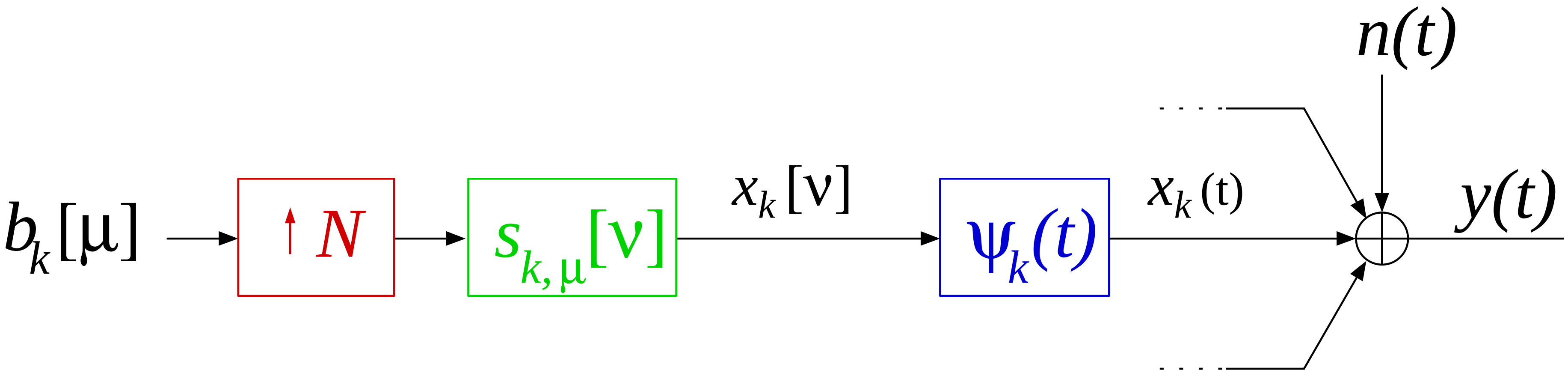


Channel is additive.

Noise is additive, white, and Gaussian distributed.

Noise is independent of X_1 and X_2.

Block Structure of Linear CDMA

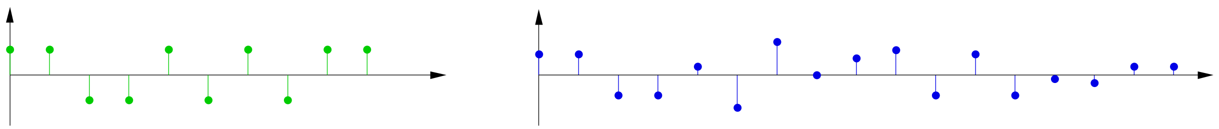


- Upsampling
- Filtering
- Pulse shaping

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:

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

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

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:

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

Analog Frequency Multiplex

Signals are orthogonal for arbitrary time shifts.

Spreading sequences of different users are the impulse responses of non-overlapping bandpass filters.

Spreading sequences are much longer than the spreading factor.

Disadvantages:

- Loss of radio spectrum due to transfer function with finite slope.
- Sensitive to multipath fading.

Frequency Hopping CDMA

Form of analog frequency multiplex which frequently changes the spreading sequence.

FH-CDMA was not invented for the purpose of multiplex, but to avoid jamming in military applications. It is the first form of CDMA reported in literature.

In 1908, Jonathan Zenneck references the invention to the German company Telefunken. FH was used by the German military in World War I.

Reinventions:

1926: Otto B. Blackwell, De Loss K. Martin, and Gilbert S. Vernam were granted U.S. Patent 1,598,673

1929: Leonard Danilewicz in Poland

1932: Willem Broertjes was granted U.S. Patent 1,869,659

1942: Hedy Kiesler Markey and George Antheil were granted U.S. Patent 2,292,387.

Frequency Hopping CDMA

Form of analog frequency multiplex which frequently changes the spreading sequence.

FH-CDMA was not invented for the purpose of multiplex, but to avoid jamming in military applications. It is the first form of CDMA reported in literature.

In 1908, Jonathan Zenneck references the invention to the German company Telefunken. FH was used by the German military in World War I.

Reinventions:

1926: Otto B. Blackwell, De Loss K. Martin, and Gilbert S. Vernam were granted U.S. Patent 1,598,673

1929: Leonard Danilewicz in Poland

1932: Willem Broertjes was granted U.S. Patent 1,869,659

1942: Hedy Kiesler Markey and George Antheil were granted U.S. Patent 2,292,387.

Hedy Kiesler Markey in 1938
born Hedwig Eva Maria Kiesler
known as actress Hedy Lamarr

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

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

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_n^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 $\boldsymbol{b}$

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

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:

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

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:

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

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.*

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

with

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

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

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

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

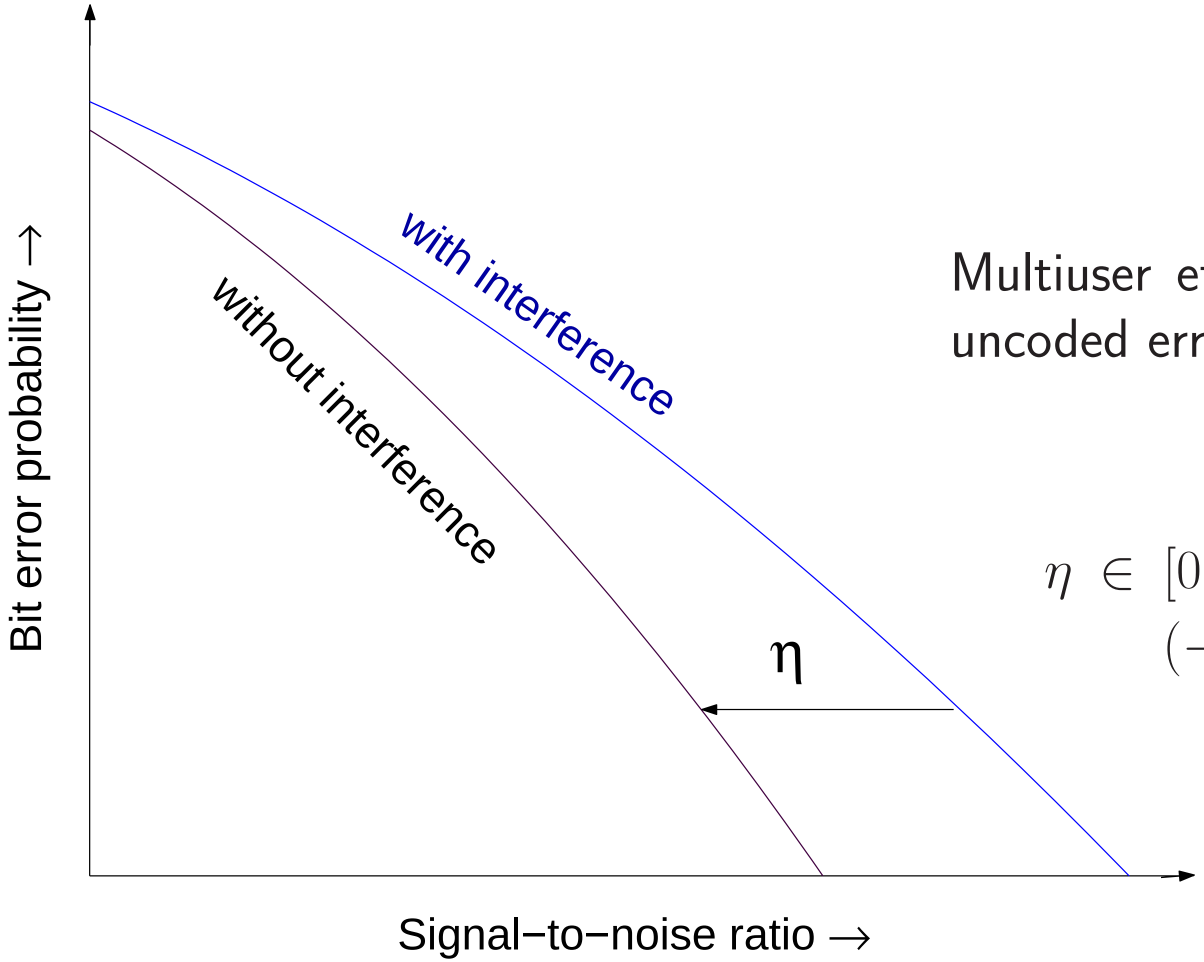
$$\mathrm{diag}\left(\boldsymbol{LRA}\right) = \mathbf{I}.$$

This leads to the detector

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

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* $P_k(\sigma_n^2, \boldsymbol{R}, \text{'det'})$ *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*

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

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

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

Sergio Verdú
born in Barcelona in 1958

Verdú introduced the notion of multiuser efficiency in 1986.

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 Decorrelator (cont'd)

Theorem 2 *Assume that the spreading sequences are i.i. complex Gaussian distributed, and normalized such that $\boldsymbol{A} = \mathbf{I}$, then the pdf of the multi–user efficiency is given by [9]*

$$\mathrm{p}_\eta(\eta) = \begin{cases} \frac{(N-3)!}{(N-K-2)!(K-2)!}\,\eta^{N-K}(1-\eta)^{K-2} & \text{for} \quad 0 \leq \eta \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Mean and variance are

$$\mathrm{E}\,\eta = 1 - \frac{K-1}{N} \qquad \mathrm{Var}\,\eta = \frac{(N-K+1)(K-1)}{N^2(N+1)}.$$

For $N \to \infty$, η becomes deterministic.

The LMMSE Detector

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

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

Multi–user efficiency is approximately

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

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,

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

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.

Chapter 2:

Random Matrix Theory

Eigen- vs. Singular Values

Singular value decomposition:

$$\forall\, \boldsymbol{S} \in \mathbb{C}^{N\times K} \;\exists\; \boldsymbol{V} \in \mathbb{C}^{N\times N}, \boldsymbol{U} \in \mathbb{C}^{K\times K}, \boldsymbol{\Sigma} \in \left(\mathbb{R}_0^+\right)^{N\times K} \quad :$$

$$\begin{aligned} &\boldsymbol{S} = \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}, \\ &\boldsymbol{V}^{\mathrm{H}}\boldsymbol{V} = \mathbf{I}, \boldsymbol{U}\boldsymbol{U}^{\mathrm{H}} = \mathbf{I}, \\ &(\boldsymbol{\Sigma})_{i,j} = 0 \forall i \neq j \end{aligned}$$

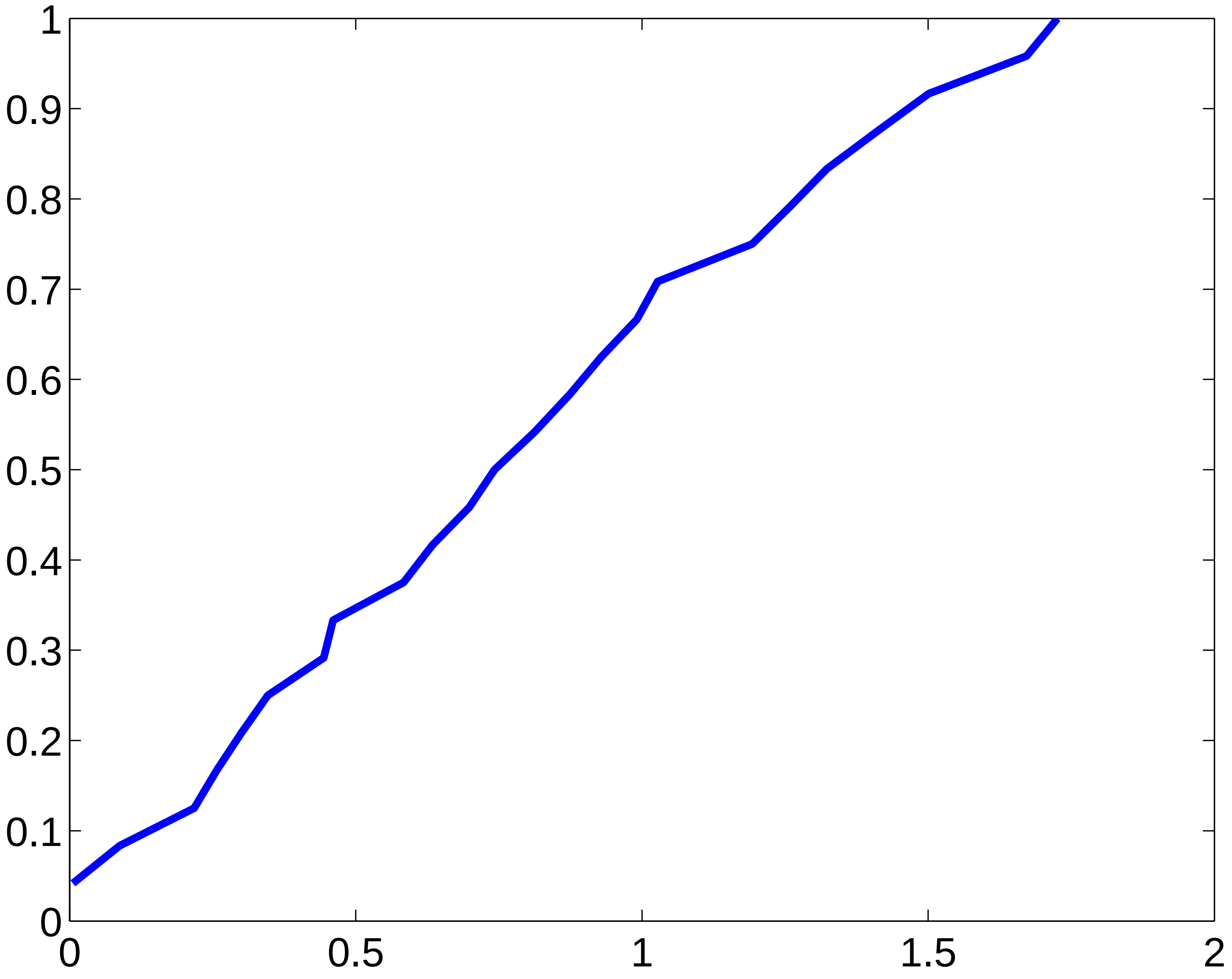
Thus,

$$\begin{aligned} \boldsymbol{S}^{\mathrm{H}}\boldsymbol{S} &= \boldsymbol{U}^{\mathrm{H}}\boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{U} \\ \boldsymbol{S}\boldsymbol{S}^{\mathrm{H}} &= \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{V}^{\mathrm{H}} \end{aligned}$$

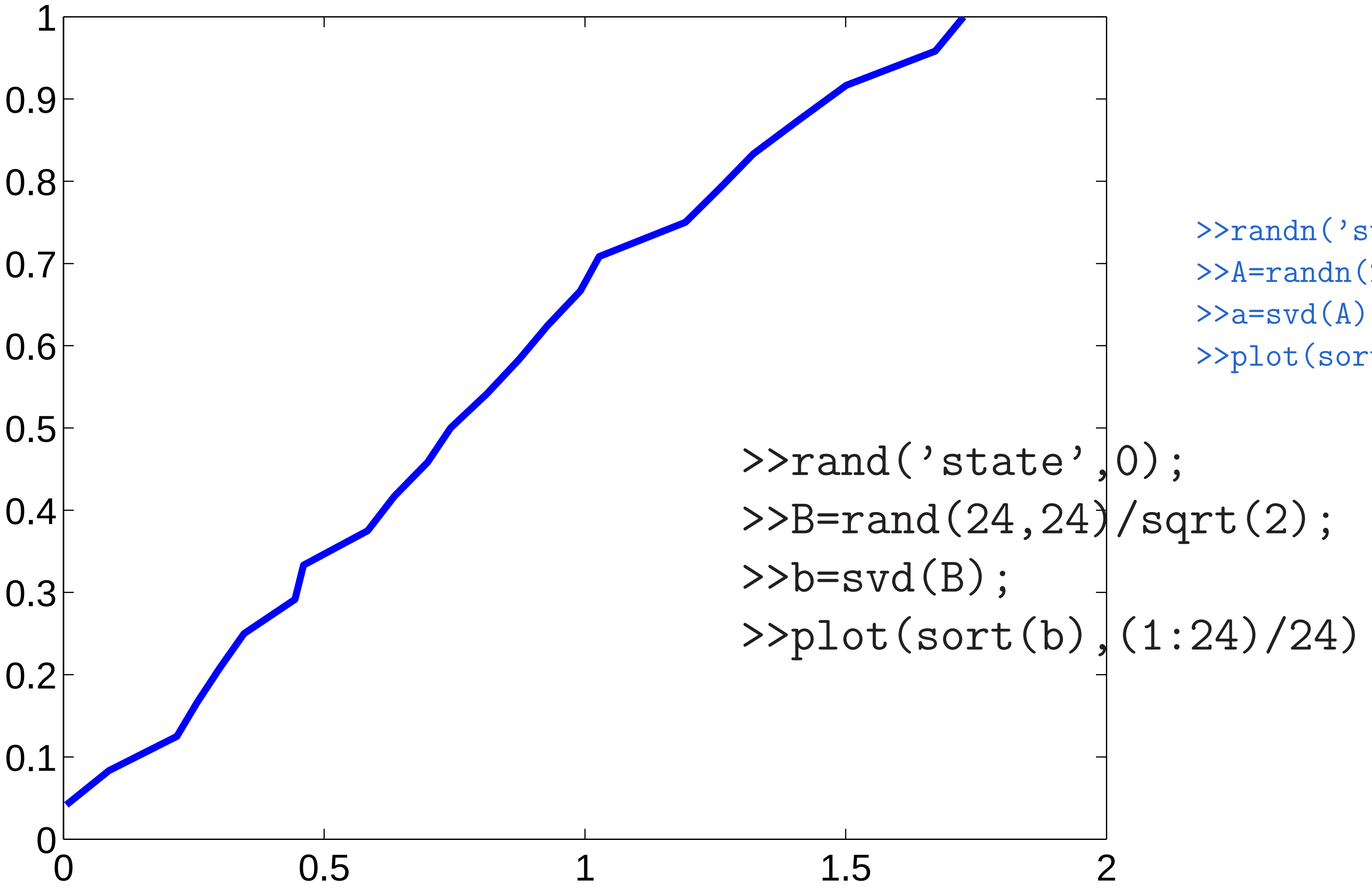
The eigenvalues of a covariance matrix are the squared singular values of the respective spreading matrix.

An Experiment with MATLAB

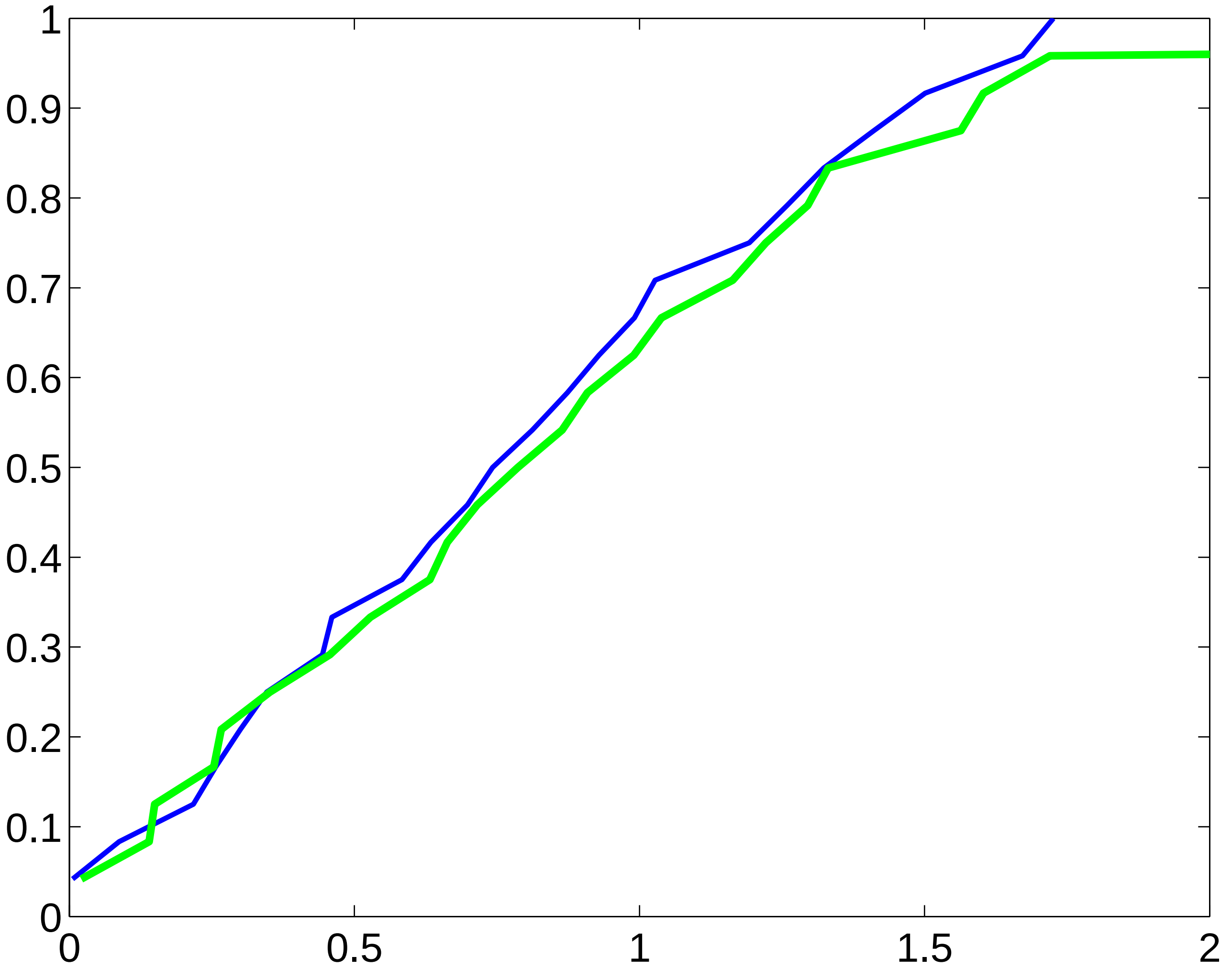
```
>>randn('state',0);
>>A=randn(24,24)/sqrt(24);
>>a=svd(A);
>>plot(sort(a),(1:24)/24)
```




```
>>randn('state',0);
>>A=randn(24,24)/sqrt(24);
>>a=svd(A);
>>plot(sort(a),(1:24)/24)
```

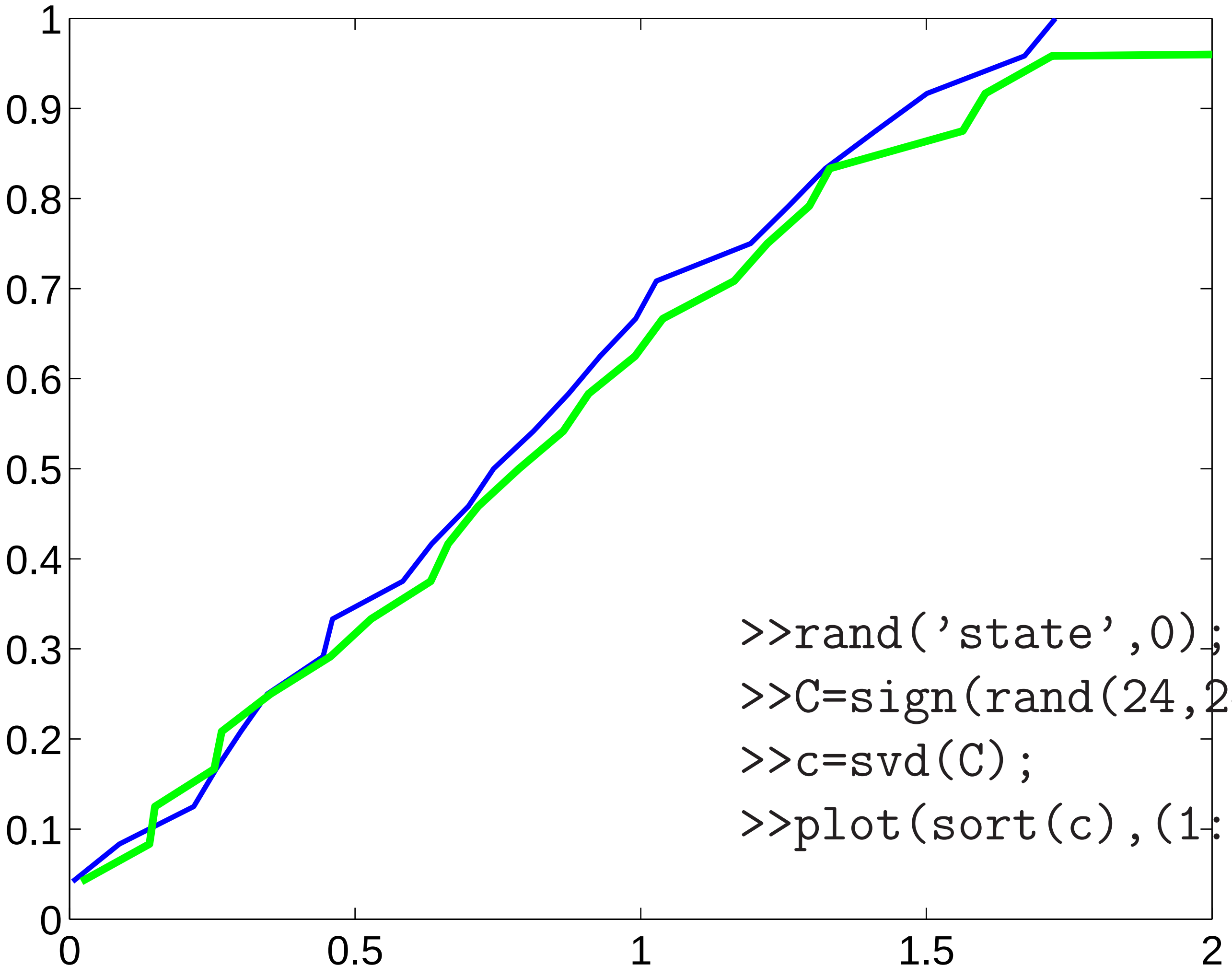



```
>>randn('state',0);
>>A=randn(24,24)/sqrt(24);
>>a=svd(A);
>>plot(sort(a),(1:24)/24)
```




```
>>randn('state',0);
>>A=randn(24,24)/sqrt(24);
>>a=svd(A);
>>plot(sort(a),(1:24)/24)
```

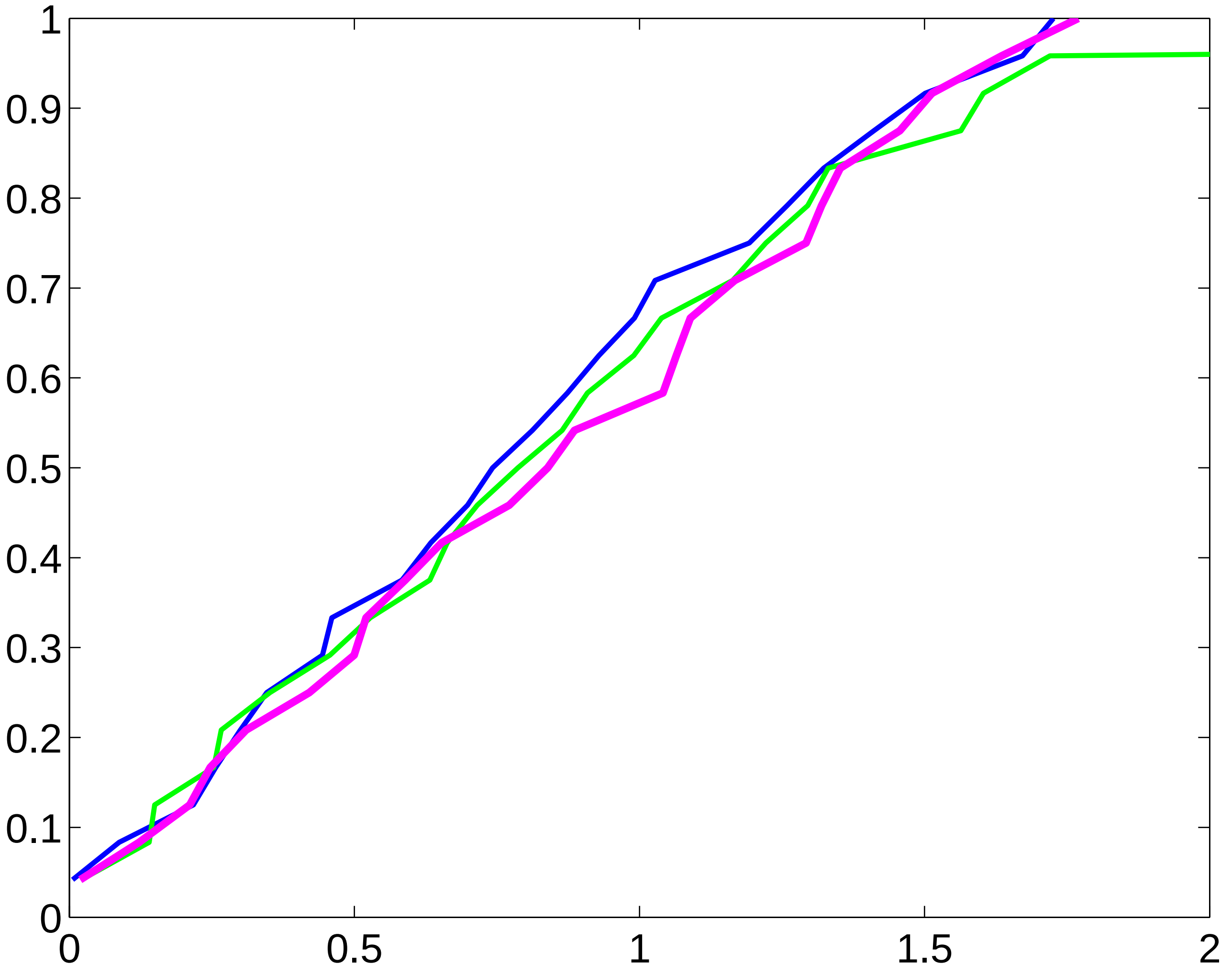
```
>>rand('state',0);
>>B=rand(24,24)/sqrt(2);
>>b=svd(B);
>>plot(sort(b),(1:24)/24)
```




```
>>randn('state',0);
>>A=randn(24,24)/sqrt(24);
>>a=svd(A);
>>plot(sort(a),(1:24)/24)
```

```
>>rand('state',0);
>>B=rand(24,24)/sqrt(2);
>>b=svd(B);
>>plot(sort(b),(1:24)/24)
```

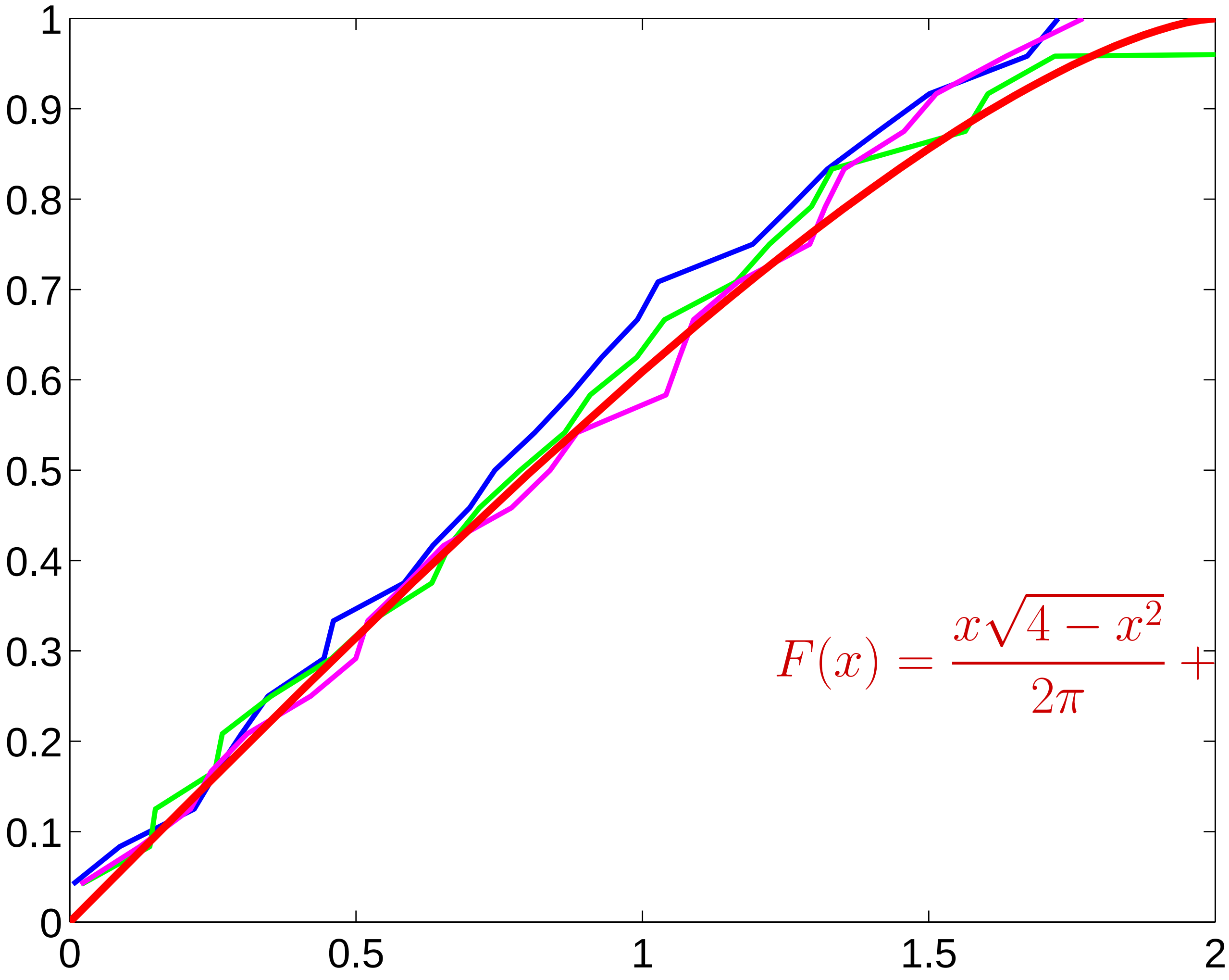
```
>>rand('state',0);
>>C=sign(rand(24,24)-.5)/sqrt(24);
>>c=svd(C);
>>plot(sort(c),(1:24)/24)
```

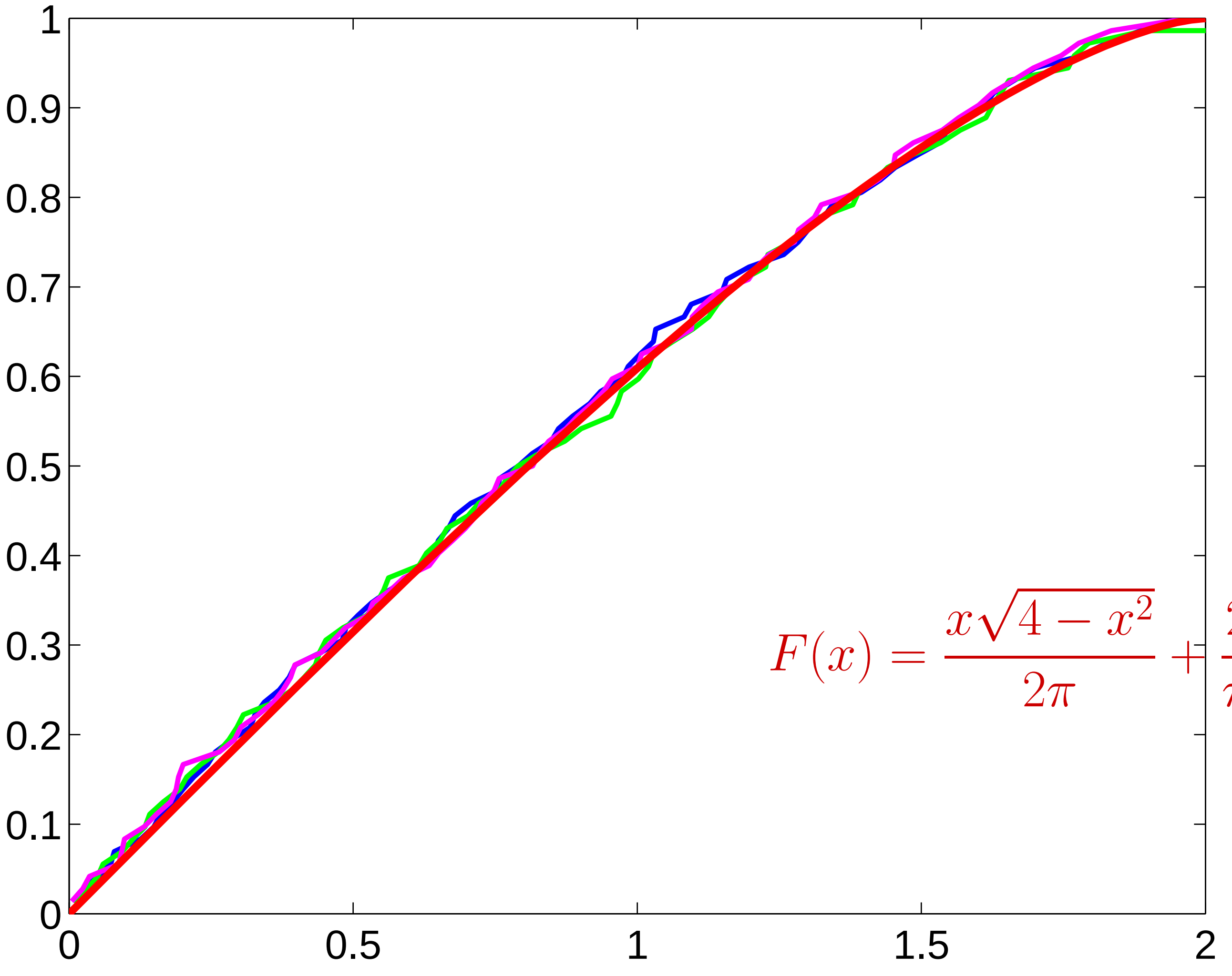

```
>>randn('state',0);
>>A=randn(24,24)/sqrt(24);
>>a=svd(A);
>>plot(sort(a),(1:24)/24)

>>rand('state',0);
>>B=rand(24,24)/sqrt(2);
>>b=svd(B);
>>plot(sort(b),(1:24)/24)

>>rand('state',0);
>>C=sign(rand(24,24)-.5)/sqrt(24);
>>c=svd(C);
>>plot(sort(c),(1:24)/24)
```



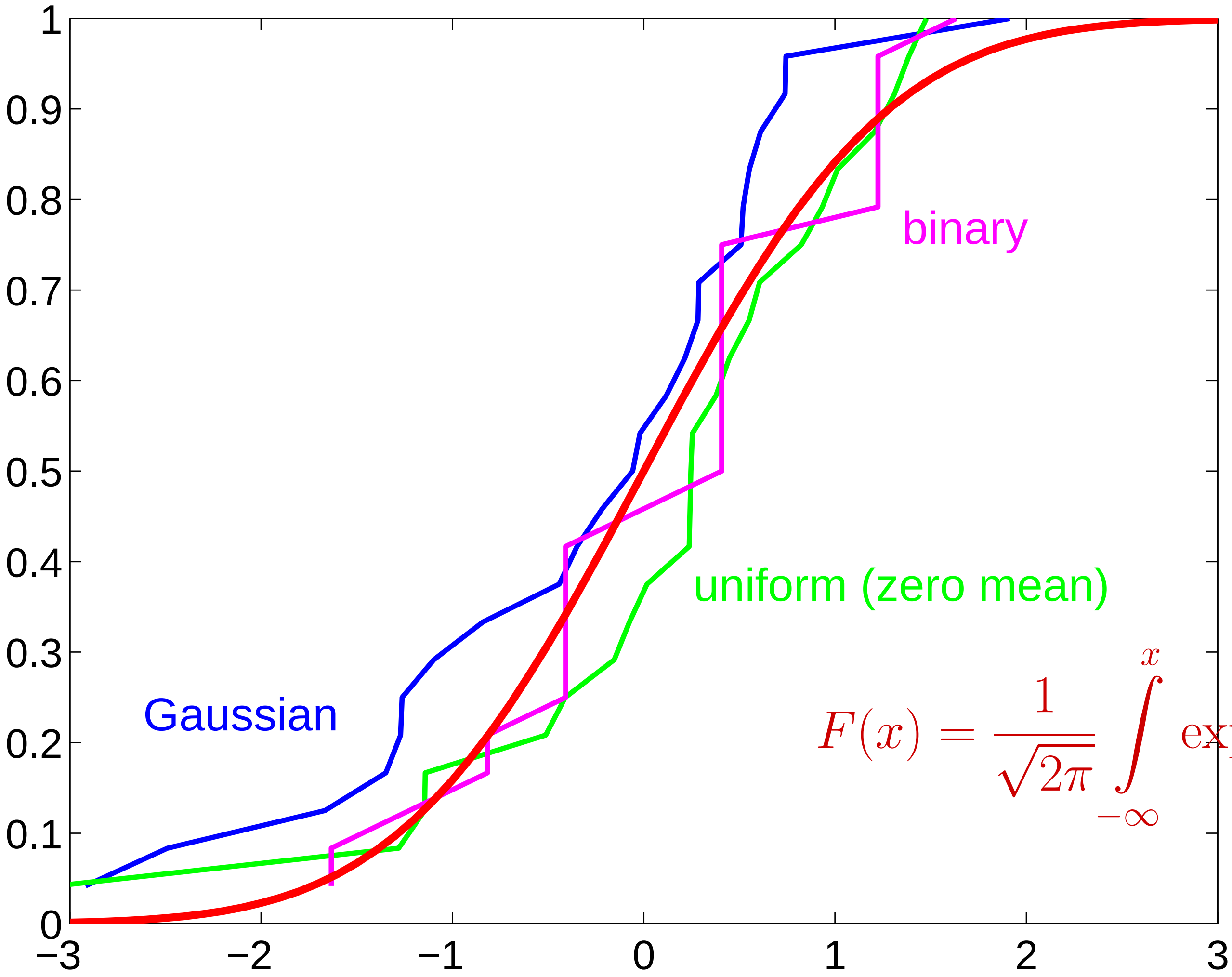
$$F(x) = \frac{x\sqrt{4-x^2}}{2\pi} + \frac{2}{\pi}\arcsin\left(\frac{x}{2}\right)$$



$$F(x) = \frac{x\sqrt{4-x^2}}{2\pi} + \frac{2}{\pi}\arcsin\left(\frac{x}{2}\right)$$

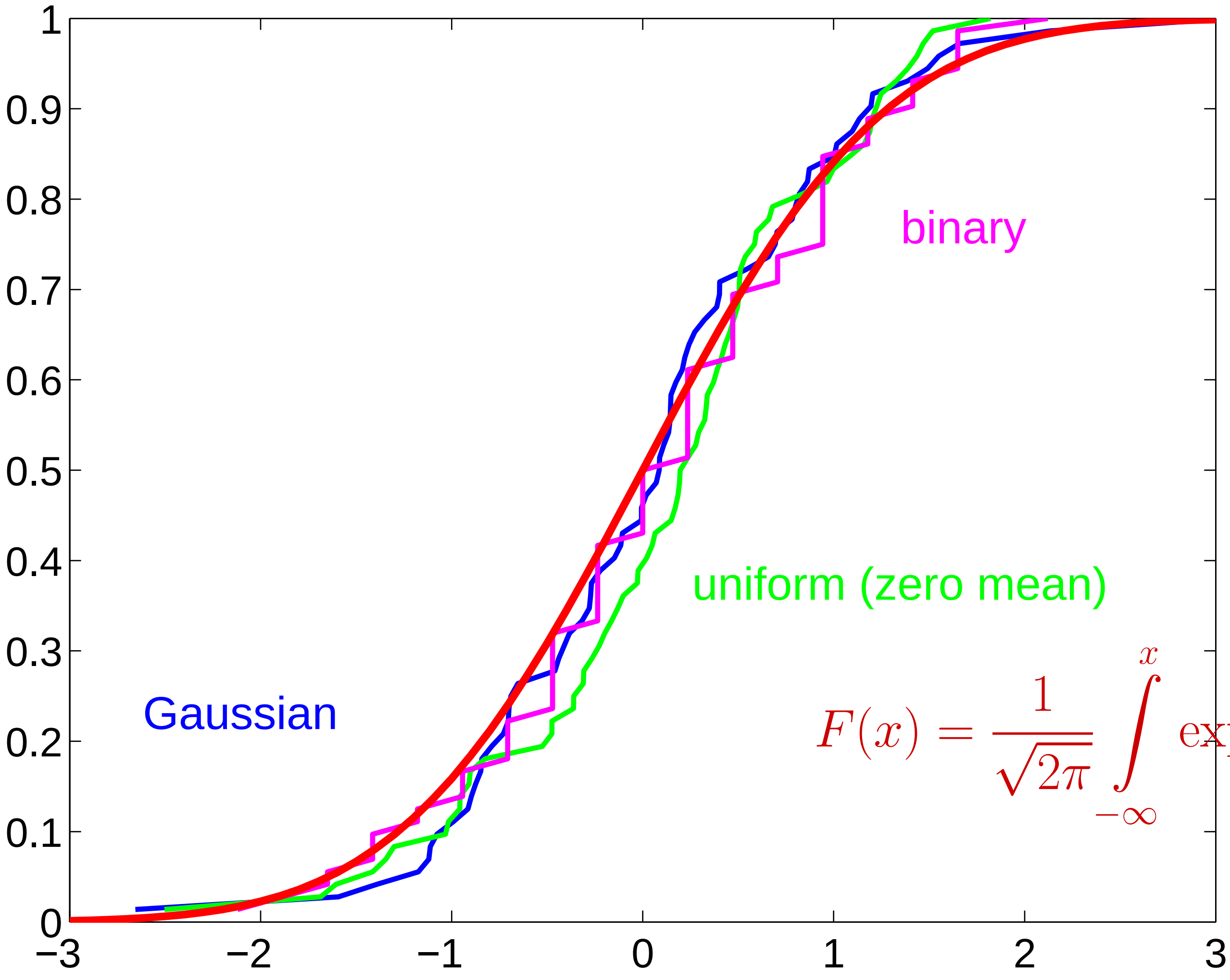
Triple matrix size to 72×72

Change `svd(·)` *to* `sum(·)`*.*



Matrix size
24×24

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{\xi^2}{2}\right) d\xi$$



Matrix size
72×72

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{\xi^2}{2}\right) \mathrm{d}\xi$$

Observations

- In both cases, the limit distribution did not depend on the distribution of the matrix entries.

- For `svd(·)` convergence is faster than for `sum(·)`.

- The limit distribution depends on the projection

$$f : \mathbb{R}^{K\times K} \mapsto \mathbb{R}^{K}.$$

The Two Theories

Random Matrix Theory (RMT) considers the limit distributions for various projection functions f and various joint distributions of the matrix elements.

Free Probability Theory (FPT) considers the large random matrix as a single random operator and develops a probability theory for non-commutative operator algebras.

The two theories are very relevant for communications engineering.

Convergence of Random Variables

Definition 7 *Given a sequence of random variables X_1, X_2, ..., we say that the sequence converges to X*

1. *in distribution (also called in law) if*

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \qquad \forall x \in \mathbb{R}$$

2. *in probability if*

$$\lim_{n\to\infty} \Pr\left\{|X_n - X| > \epsilon\right\} = 0 \qquad \forall \epsilon > 0,$$

3. *in $r^{\rm th}$ mean if*

$$\lim_{n\to\infty} \mathrm{E}\left\{|X_n - X|^r\right\} = 0,$$

4. *almost surely (also called with probability 1) if*

$$\Pr\left\{\lim_{n\to\infty} X_n = X\right\} = 1.$$

Convergence of Random Variables (cont'd)

Theorem 3 *Almost sure convergence implies convergence in probability.*

Theorem 4 *Convergence in any mean $r > s$ implies convergence in any mean $s \geq 1$.*

Theorem 5 *Convergence in any positive mean implies convergence in probability.*

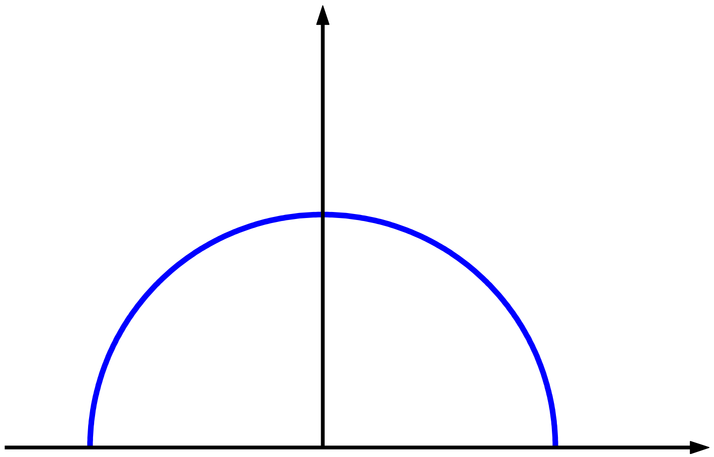
Theorem 6 *Convergence in probability implies convergence in distribution.*

Semi-Circle Law *(Wigner 1955)*

For a $K \times K$ random matrix $\boldsymbol{H}$ with i.i.d. zero mean elements of variance $\frac{1}{K}$, the empirical distribution of the eigenvalues of $\frac{1}{\sqrt{2}}(\boldsymbol{H} + \boldsymbol{H}^{\mathrm{H}})$ converges almost surely to a deterministic limit distribution with density

$$f(x) = \frac{1}{2\pi}\sqrt{4 - x^2} \qquad x \in (-2; +2)$$

as $K \to \infty$.



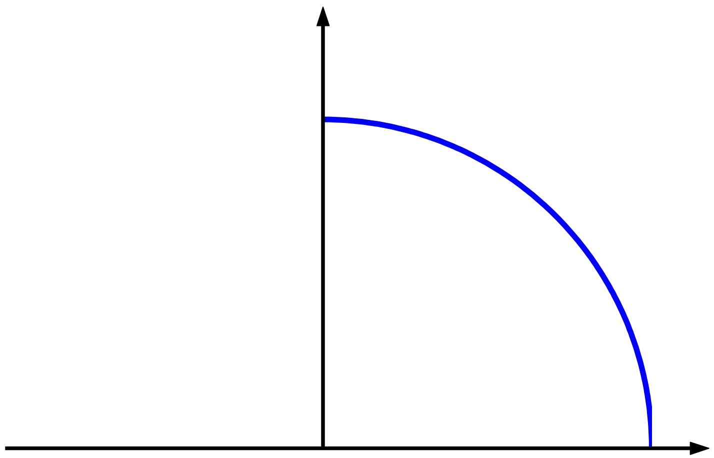
Eugene Paul Wigner
born in Budapest in 1902.

Quarter Circle Law

For a $K \times K$ random matrix with i.i.d. elements of variance $\frac{1}{K}$, the empirical distribution of the singular values converges almost surely to a deterministic limit distribution with density

$$f(x) = \frac{1}{\pi}\sqrt{4 - x^2} \qquad x \in [0; 2)$$

as $K \to \infty$.



Full Circle Law

For a $K \times K$ random matrix with i.i.d. zero-mean elements of variance $\frac{1}{K}$, the empirical distribution of the eigenvalues converges almost surely to a deterministic limit distribution with density

$$f(z) = \frac{1}{\pi} \qquad |z| < 1$$

as $K \to \infty$.

Uniform distribution on the complex unit disc.

Haar Distribution

For a $K \times K$ random matrix $\boldsymbol{H}$ with i.i.d. zero-mean Gaussian elements of finite variance, the empirical distribution of the eigenvalues of $\boldsymbol{T} = \boldsymbol{H}(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H})^{-\frac{1}{2}}$ converges almost surely to a deterministic limit distribution with density

$$f(z) = \frac{1}{2\pi} \qquad |z| = 1$$

as $K \to \infty$.

Uniform distribution on the complex unit circle.

The matrix $\boldsymbol{T}$ is uniformly distributed on the set of unitary random matrices.

Inverse Semi-Circle Law

Let the $K \times K$ random matrix $\boldsymbol{H}$ be composed of i.i.d. zero-mean Gaussian elements of finite variance, and define the matrix

$$\boldsymbol{T} = \boldsymbol{H}(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H})^{-\frac{1}{2}}.$$

Moreover, let

$$\boldsymbol{Y} = \boldsymbol{T} + \boldsymbol{T}^{\mathrm{H}}.$$

Then, the empirical distribution of the eigenvalues of $\boldsymbol{Y}$ converges almost surely to a non-random distribution function as $K \to \infty$ whose density is given by

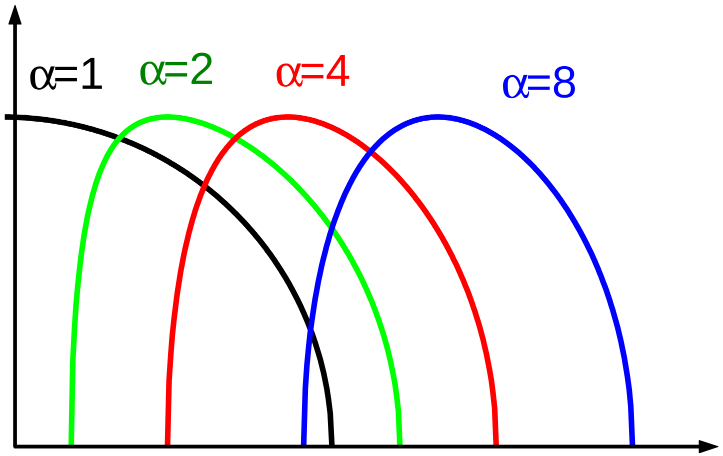
$$f(x) = \frac{1}{\pi}\frac{1}{\sqrt{4-x^2}} \qquad x \in (-2;+2)$$

Deformed Quarter Circle Law

For an $N \times K$, $N < K$ random matrix with i.i.d. elements of variance $\frac{1}{N}$, the empirical distribution of the singular values converges almost surely to a deterministic limit distribution with density

$$f(x) = \frac{\sqrt{4\alpha - (x^2 - 1 - \alpha)^2}}{\pi x} \qquad x \in (\sqrt{\alpha} - 1; \sqrt{\alpha} + 1)$$

as $K = \alpha N \to \infty$.



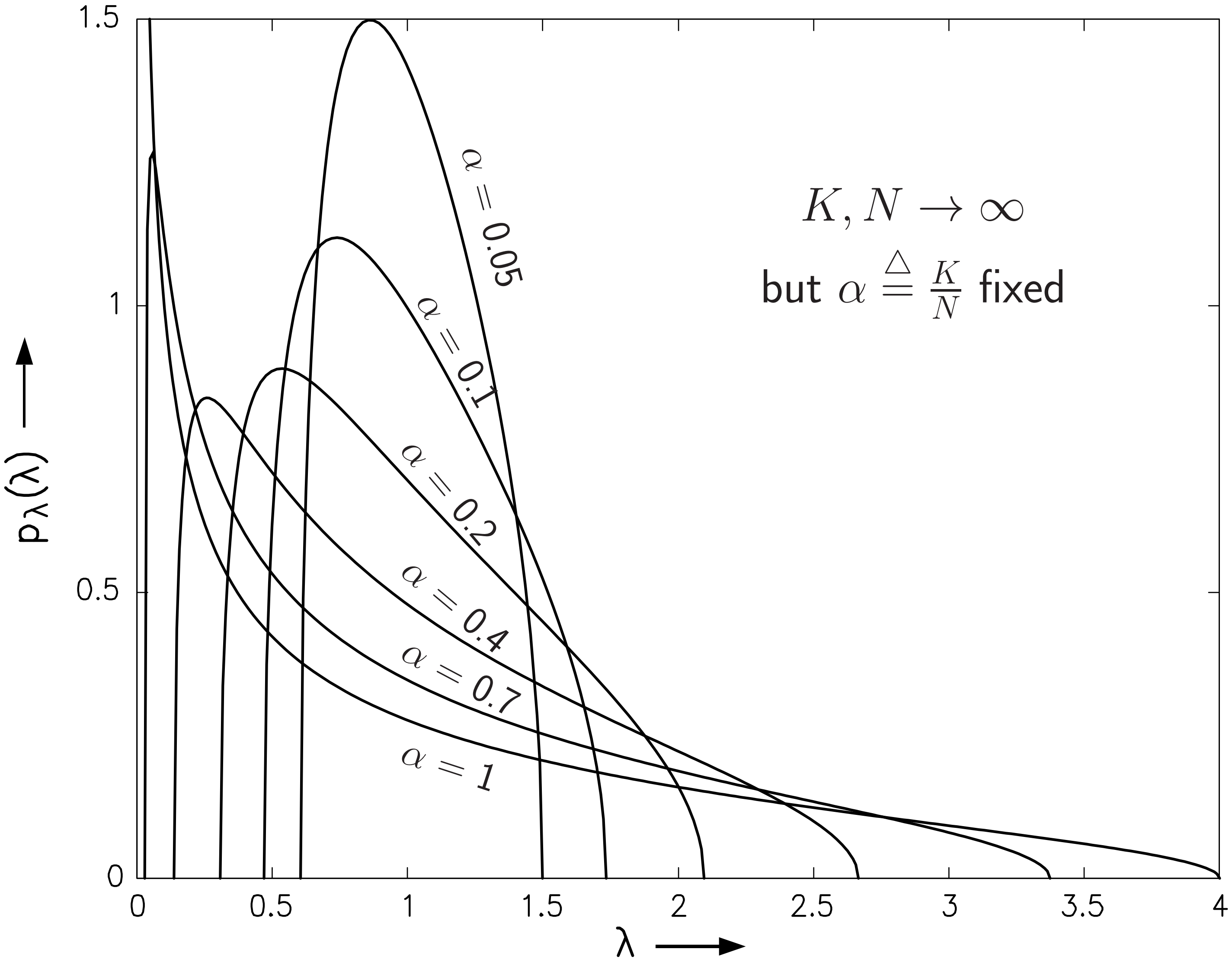
Deformed Quarter Circle Law (eigenvalues)

As $K = \alpha N \to \infty$:

$$\mathrm{p}_\lambda(x) = \begin{cases} \dfrac{\sqrt{\left(x - (1-\sqrt{\alpha})^2\right)\left((1+\sqrt{\alpha})^2 - x\right)}}{2\pi\alpha x} & \text{for } (1-\sqrt{\alpha})^2 < x < (1+\sqrt{\alpha})^2 \\ \left(1 - \dfrac{1}{\alpha}\right)\begin{cases} 0 & \text{for } \alpha \le 1 \\ \delta(x) & \text{for } \alpha > 1 \end{cases} & \text{otherwise} \end{cases}$$

This appears to be inconsistent with page 52. What is different here?

Deformed Quarter Circle Law (eigenvalues)



Convergence Properties of Eigenvectors

Let $\boldsymbol{H}$ be an $N \times K$ random matrix with i.i.d. real-valued random entries with zero mean and all positive moments bounded from above. Let the orthogonal matrix $\boldsymbol{U}$ be defined by the eigenvalue decomposition

$$\boldsymbol{U}^{\mathrm{T}} \boldsymbol{\Lambda} \boldsymbol{U} = \boldsymbol{H}^{\mathrm{T}} \boldsymbol{H}.$$

Let $\boldsymbol{x} \in \mathbb{R}^N$ be an arbitrary vector with unit Euclidean norm and the random vector $\boldsymbol{y} = [y_1, \ldots, y_N]^{\mathrm{T}}$ be defined as

$$\boldsymbol{y} = \boldsymbol{U} \boldsymbol{x}.$$

Then, as $K = \alpha N \to \infty$,

$$\sum_{k=1}^{\lceil tN \rceil} y_k^2 \longrightarrow t$$

almost surely for every $t \in [0; 1]$.

This result is like a law of large numbers for the components of any linear combination of the components of the eigenvectors of $\boldsymbol{H}^{\mathrm{T}} \boldsymbol{H}$.

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:

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

The Moments Problem

In 1894, Stieltjes looked at the problem of finding a distribution given its moments

$$m_n = \int x^n \mathrm{dF}(x).$$

He introduced the power series

$$G(s) = -\sum_{n=0}^{k} m_n s^{-n-1} + \mathrm{o}\left(s^{-k-1}\right)$$

of the complex variable s and found the Stieltjes inversion formula as the solution to the moments problem.

Thomas Joannes Stieltjes
born in Zwolle in 1856

Some Stieltjes Transforms

identity : $\mathrm{G}(s) = \dfrac{1}{1-s}$

semi-circle : $\mathrm{G}(s) = \dfrac{s}{2}\sqrt{1-\dfrac{4}{s^2}} - \dfrac{s}{2}$

quarter circle : $\mathrm{G}(s) = \dfrac{2\sqrt{4-s^2}}{\pi} \ln\left(\dfrac{2+\sqrt{4-s^2}}{-s}\right) - \dfrac{s}{2} - \dfrac{2}{\pi}$

$(\text{quarter circle})^2$: $\mathrm{G}(s) = \dfrac{1}{2}\sqrt{1-\dfrac{4}{s}} - \dfrac{1}{2}$

$(\text{def. quarter circle})^2$: $\mathrm{G}(s) = \sqrt{\dfrac{(1-\alpha)^2}{4s^2} - \dfrac{1+\alpha}{2s} + \dfrac{1}{4}} - \dfrac{1}{2} - \dfrac{1-\alpha}{2s}$

inverse semi-circle : $\mathrm{G}(s) = \dfrac{1}{\sqrt{s^2-4}}$

projector : $\mathrm{G}(s) = \dfrac{\alpha}{1-s} + \dfrac{1-\alpha}{-s} = \dfrac{1-\alpha-s}{s^2-s}$

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

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

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.

Products of Random Matrices

Let the random matrix $\boldsymbol{H}$ fulfill the same conditions as needed for the deformed quarter circle law. Moreover, let $\boldsymbol{X} = \boldsymbol{X}^{\mathrm{H}}$ be an $N \times N$ Hermitian matrix, independent of $\boldsymbol{H}$, with an empirical eigenvalue distribution converging almost surely in distribution to a distribution function $\mathrm{P}_{\boldsymbol{X}}(x)$ as $N \to \infty$. Then, almost surely, the eigenvalue distribution of the matrix product

$$\boldsymbol{P} = \boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{X}$$

converges in distribution, as $K, N \to \infty$, but $\alpha = K/N$ fixed, to a nonrandom distribution function whose Stieltjes transform satisfies

$$\mathrm{G}_{\boldsymbol{P}}(s) = \int \frac{\mathrm{dP}_{\boldsymbol{X}}(x)}{x\big(1 - \alpha - \alpha s \mathrm{G}_{\boldsymbol{P}}(s)\big) - s}$$

for $\Im s > 0$.

Sums of Random Matrices

Let the random matrix $\boldsymbol{H}$ fulfill the same conditions as needed for the deformed quarter circle law. Let $\boldsymbol{X} = \boldsymbol{X}^{\mathrm{H}}$ be an $N \times N$ Hermitian matrix with an eigenvalue distribution function converging weakly to $\mathrm{P}_{\boldsymbol{X}}(x)$ almost surely. Let $\boldsymbol{Y} = \mathrm{diag}\,(y_1, \ldots, y_K)$ be a $K \times K$ diagonal matrix and the empirical distribution function of $\{y_1, \ldots, y_K\} \in \mathbb{R}^K$ converge almost surely in distribution to a probability distribution function $\mathrm{P}_{\boldsymbol{Y}}(x)$ as $K \to \infty$. Moreover, let the matrices $\boldsymbol{H}, \boldsymbol{X}, \boldsymbol{Y}$ be jointly independent. Then, almost surely, the empirical eigenvalue distribution of the random matrix

$$\boldsymbol{S} = \boldsymbol{X} + \boldsymbol{H}\boldsymbol{Y}\boldsymbol{H}^{\mathrm{H}}$$

converges weakly, as $K, N \to \infty$, but $\alpha = K/N$ fixed, to the unique nonrandom distribution function whose Stieltjes transform satisfies

$$\mathrm{G}_{\boldsymbol{S}}(s) = \mathrm{G}_{\boldsymbol{X}}\left(s - \alpha \int \frac{y\,\mathrm{dP}_{\boldsymbol{Y}}(y)}{1 + y\,\mathrm{G}_{\boldsymbol{S}}(s)}\right)$$

for $\Im s > 0$.

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

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

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.

Girko's Law

Let the $N \times K$ random matrix $\boldsymbol{H}$ be composed of independent entries $(\boldsymbol{H})_{ck}$ with zero-mean and variances w_{ck}^2/N such that all w_{ck}^2 are uniformly bounded from above. Assume that the empirical joint distribution of variances $w^2 : [0,1]^2 \mapsto \mathbb{R}$ defined by $w^2(x,y) = w_{ck}^2$ for c, k satisfying

$$\frac{c}{N} \leq x \leq \frac{c+1}{N} \qquad \text{and} \qquad \frac{k}{K} \leq y \leq \frac{k+1}{K}$$

converges to a bounded joint limit distribution $w^2(x,y)$ as $K = \alpha N \to \infty$. Then, for each $a, b \in [0,1], a < b$, and $\Im(s) > 0$

$$\frac{1}{N} \sum_{c=\lceil aN \rceil}^{\lceil bN \rceil} \left(\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}} - s\mathbf{I} \right)_{cc}^{-1} \longrightarrow \int_a^b g(x,s) \mathrm{d}x$$

where convergence is in probability and the function $g(x,s)$ is a chip-dependent Stieltjes transform.

Girko's Law (cont'd)

The function $g(x,s)$ satisfies the fixed point equation

$$g(x,s) = \left[-s + \alpha \int_0^1 \frac{w^2(x,y)\,\mathrm{d}y}{1 + \int_0^1 g(x',s) w^2(x',y)\,\mathrm{d}x'} \right]^{-1}$$

for every $x \in [0,1]$. The solution always exists and is unique in the class of functions $g(x,s) \geq 0$, analytic for $\Im(s) > 0$ and continuous on $x \in [0,1]$.

Moreover, almost surely, the empirical eigenvalue distribution of $\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}$ converges weakly to a limiting distribution whose Stieltjes transform is given by

$$G_{\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}}(s) = \int_0^1 g(x,s)\,\mathrm{d}x.$$

Girko has also studied more general cases of random matrices with statistically depended entries that are relevant in communications engineering, but they exceed the introductory scope of this course.

Unitary Invariance

Definition 8 *A Hermitian random matrix $\boldsymbol{X}$ is called unitarily invariant, if the joint distributions of the entries of $\boldsymbol{X}$ and $\boldsymbol{U}\boldsymbol{X}\boldsymbol{U}^{\mathrm{H}}$ are identical for any unitary matrix $\boldsymbol{U}$ that is independent of $\boldsymbol{X}$.*

Definition 9 *A rectangular random matrix $\boldsymbol{X}$ is called bi-unitarily invariant, if the joint distributions of the entries of $\boldsymbol{X}$ and $\boldsymbol{U}\boldsymbol{X}\boldsymbol{V}$ are identical for any unitary matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ that are independent of $\boldsymbol{X}$.*

Haar matrices and i.i.d. Gaussian matrices are bi-unitarily invariant.

Lemma 2 *If $\boldsymbol{X}$ is bi-unitarily invariant, $\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}$ is unitarily invariant.*

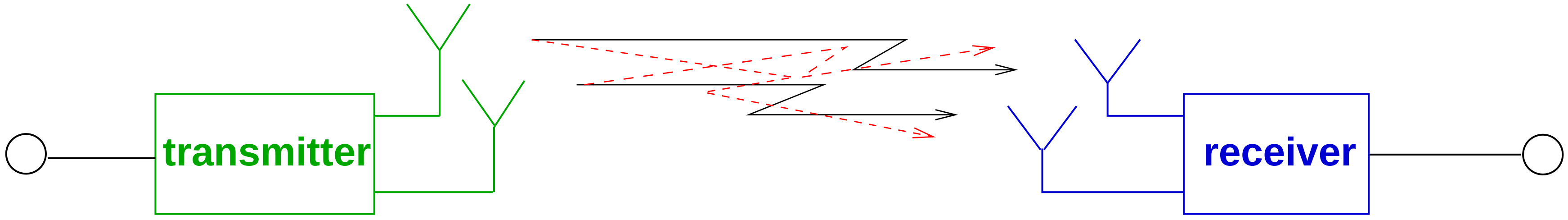
Chapter 3:

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.

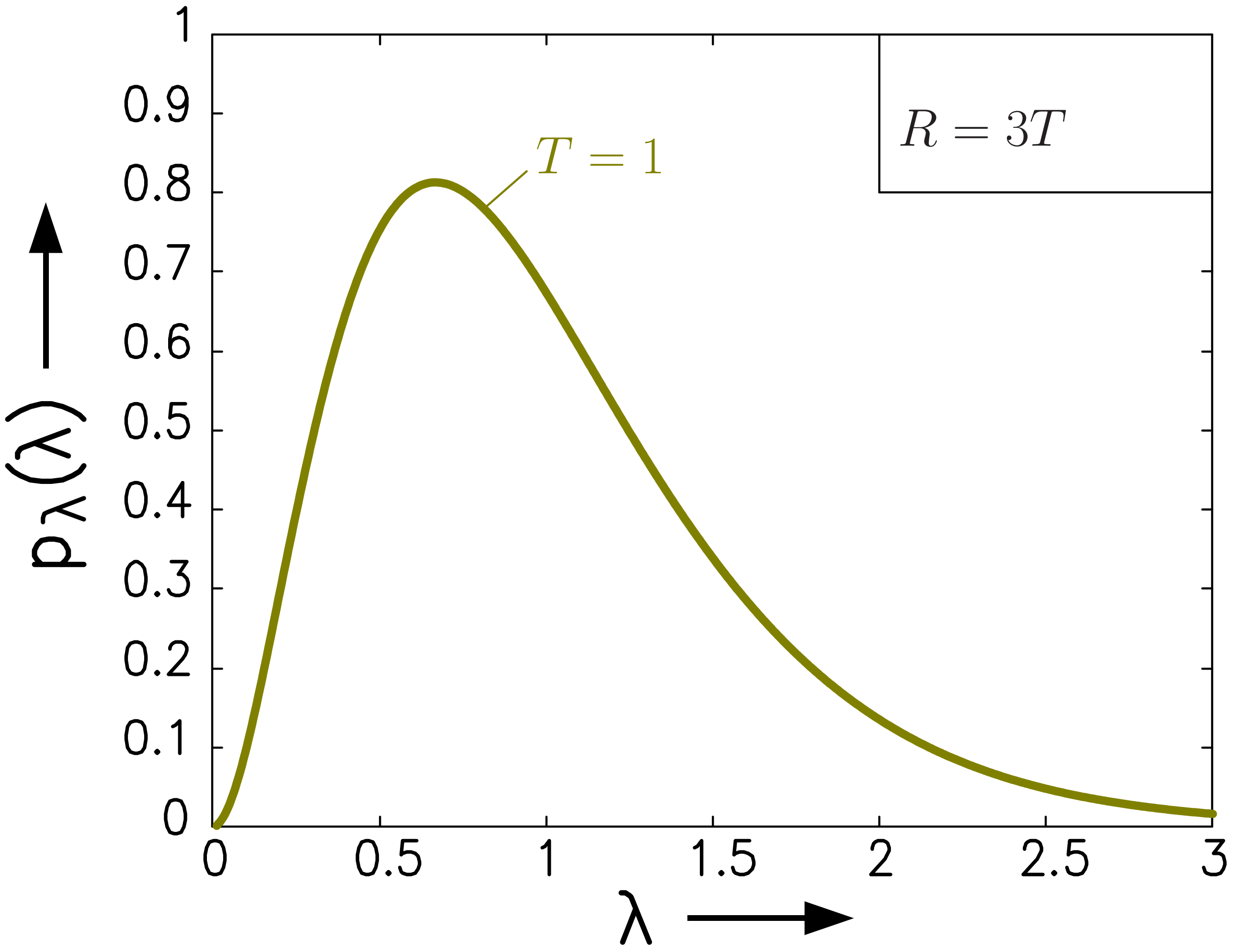
Then, the eigenvalues of $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}$ are distributed as

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

with the Laguerre polynomials

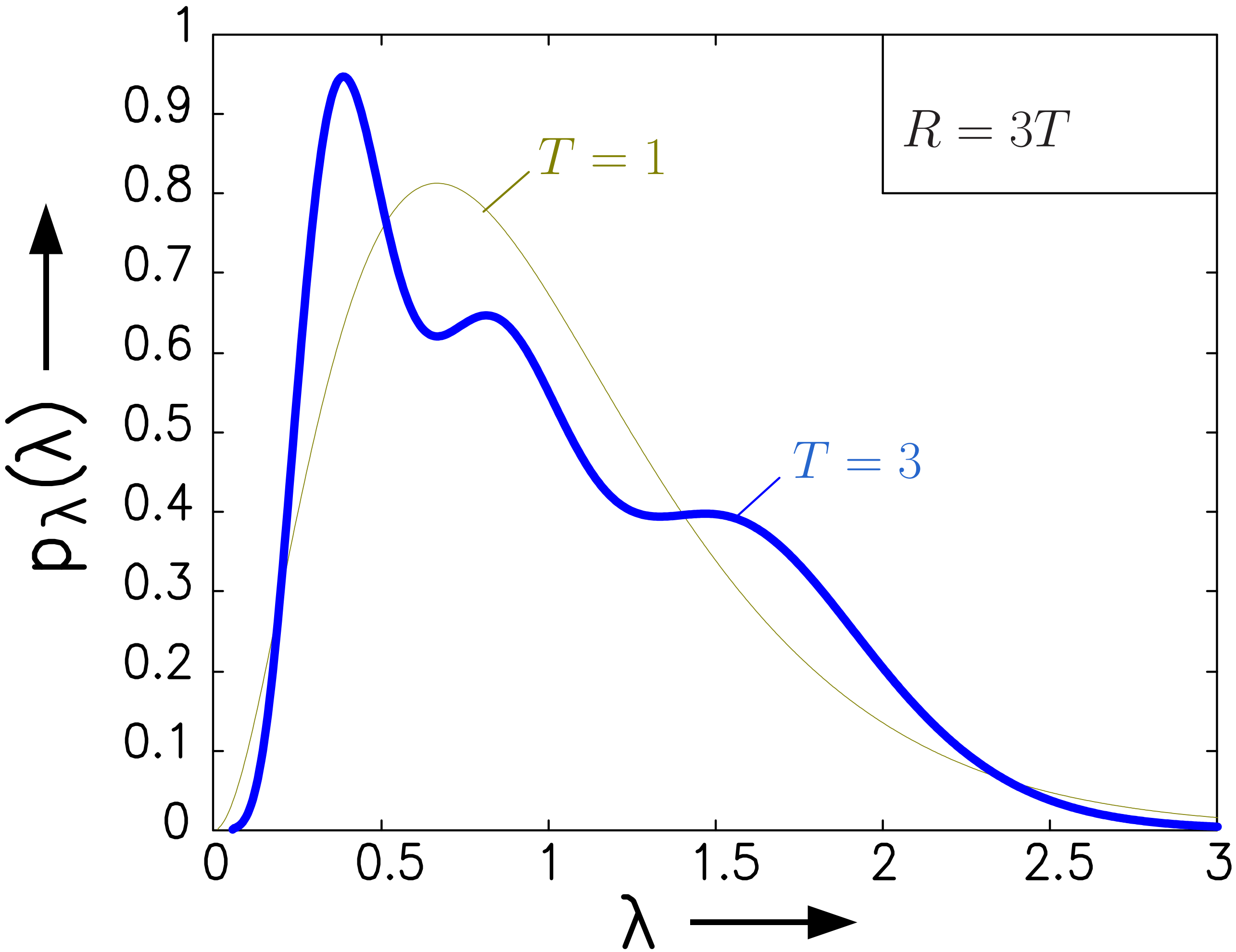
$$\mathrm{L}_a^{(b)}(x) \triangleq \frac{\mathrm{e}^x}{a!x^b} \frac{\mathrm{d}^a}{\mathrm{d}x^a}\left(x^{a+b}\,\mathrm{e}^{-x}\right)$$

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



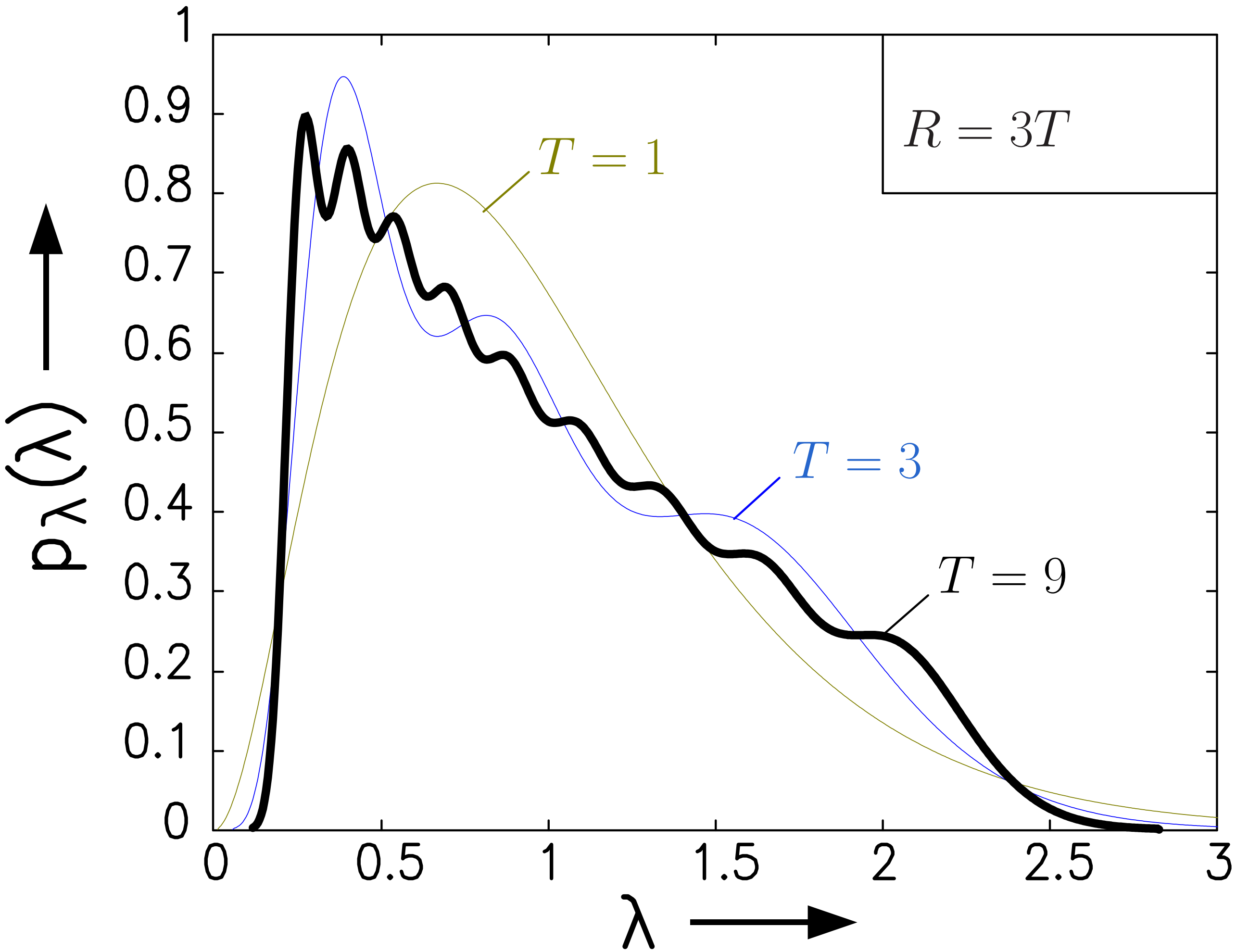
Eigenvalues of $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}$ for i.i.d. entries in the $R \times T$ matrix $\boldsymbol{H}$.

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



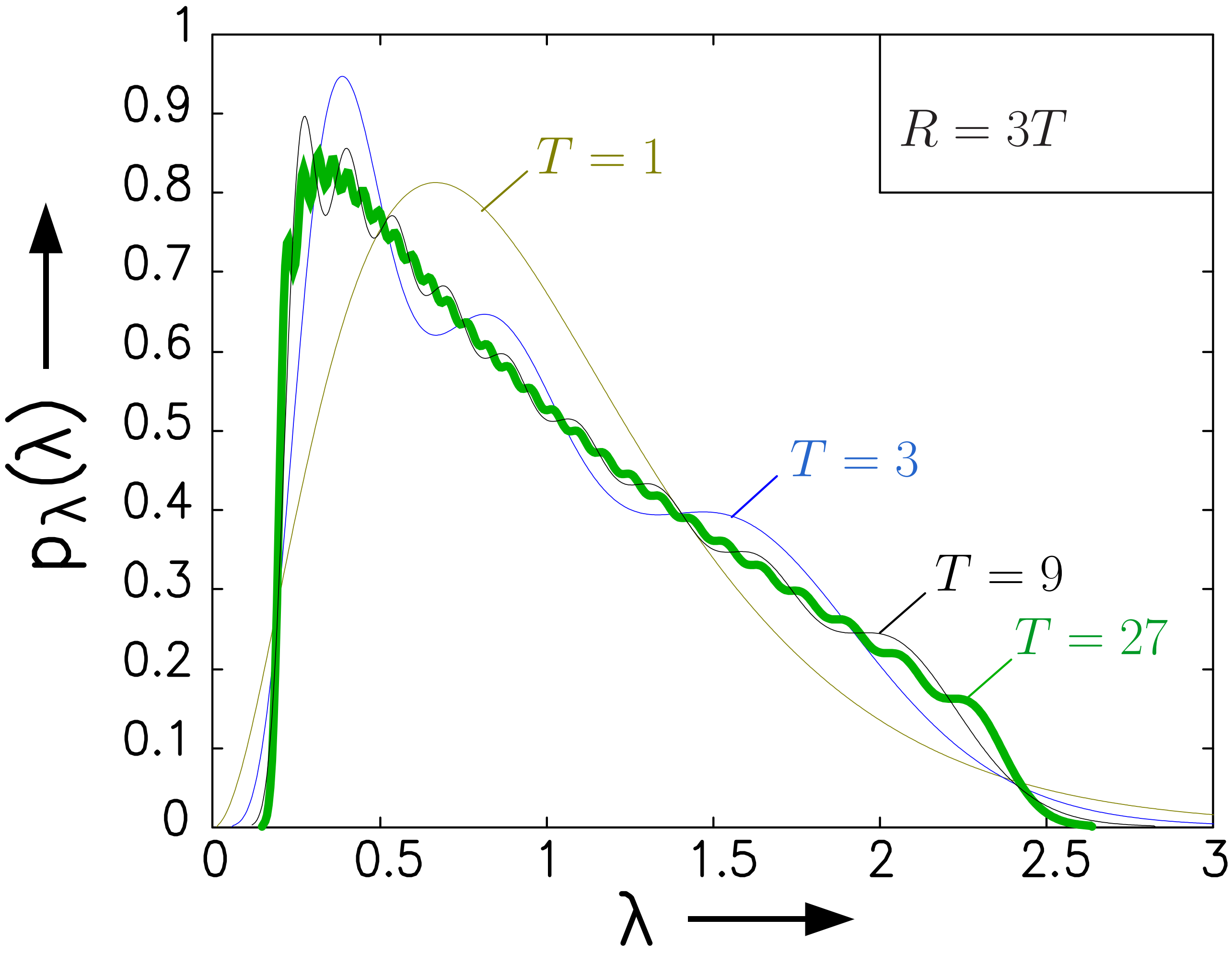
Eigenvalues of $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}$ for i.i.d. entries in the $R \times T$ matrix $\boldsymbol{H}$.

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



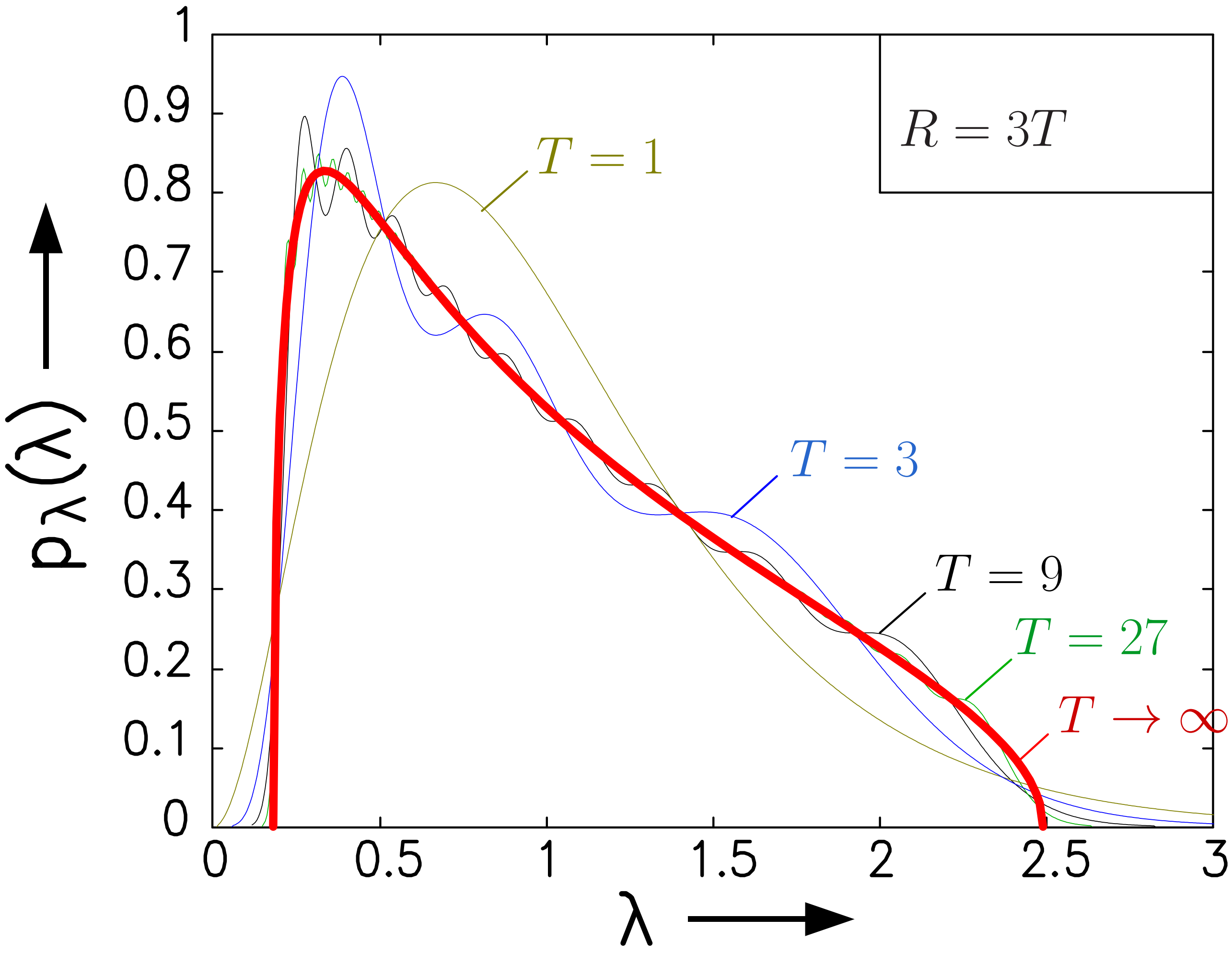
Eigenvalues of $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}$ for i.i.d. entries in the $R \times T$ matrix $\boldsymbol{H}$.

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*

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

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*

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

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

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*

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

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

Large system approximation:

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

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

Approximation by power series:

Cayley–Hamilton theorem yields:

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

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:

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

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:

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

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

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*

$$\eta_{\mathrm{WLPIC},D+1} = \cfrac{1}{1 + \cfrac{\alpha}{\sigma_n^2 + \eta_{\mathrm{WLPIC},D}}}$$

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

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

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

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.

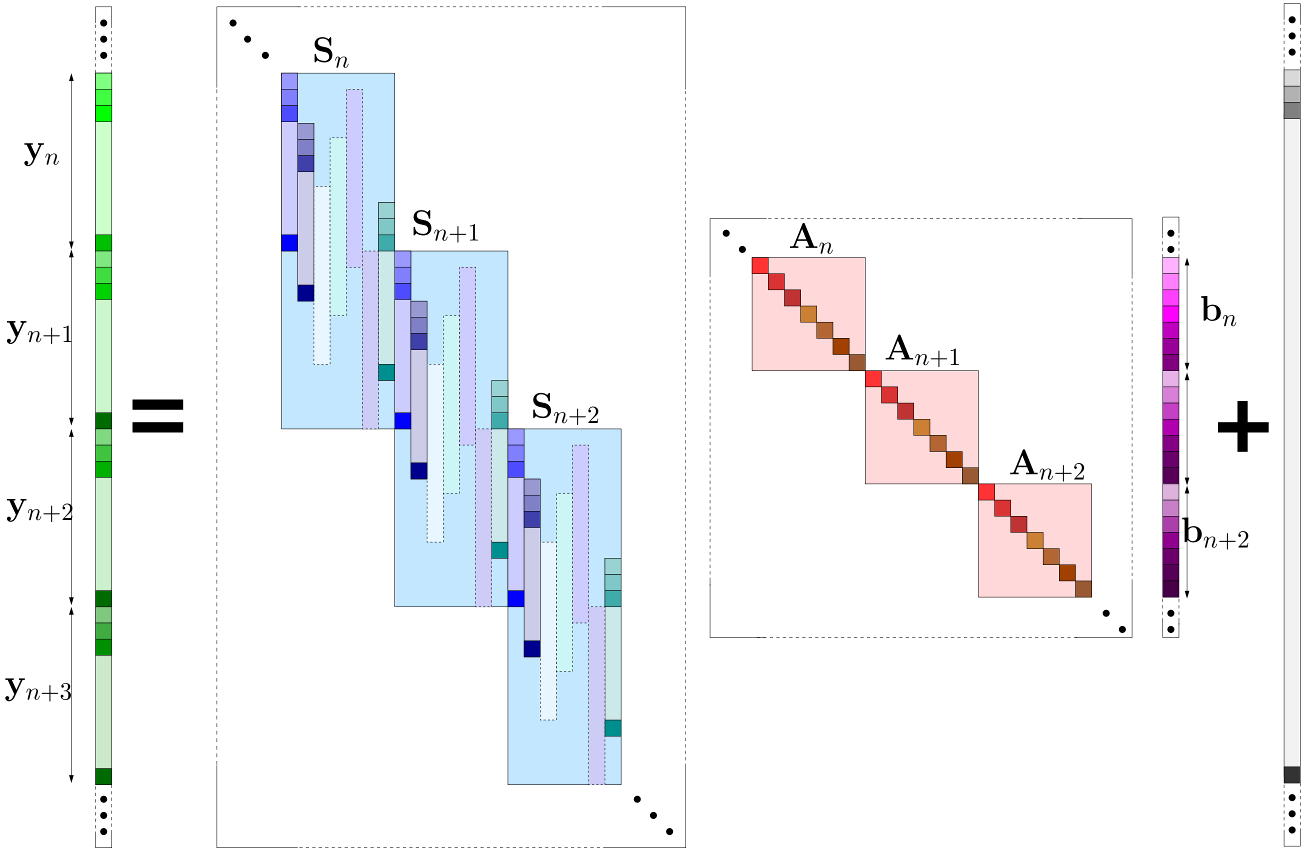
$$\mathrm{E}\left\{\boldsymbol{b}\boldsymbol{b}^{\mathrm{H}}\right\} = \mathbf{I} \otimes \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

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.

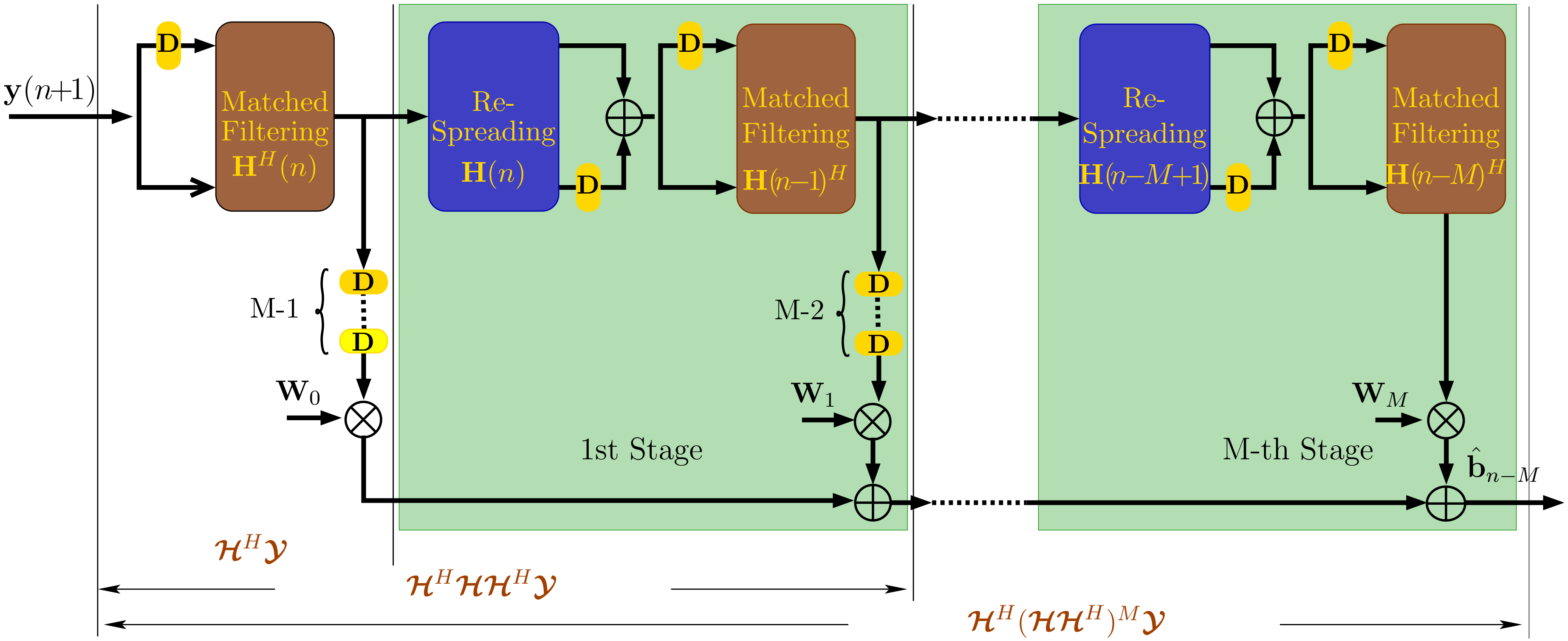
Asynchronous Users



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.

Chapter 5:

Free Probability Theory

Classical vs. Free Probability

Random matrix is ensemble	Random matrix is a sample
Multiplication is commutative	Multiplication is *not* commutative

Non-commutative joint moments:

$$\mathrm{E}\left\{\boldsymbol{A}^2\boldsymbol{B}^2\right\} \neq \mathrm{E}\left\{\boldsymbol{ABAB}\right\}$$

Statistical independence cannot be defined without respect to the elements of the matrix.

Independence vs. Freeness

Freeness of Two Random Matrices

Define for any $K \times K$ matrix $\boldsymbol{X}$

$$\mathrm{Tr}\left(\boldsymbol{X}\right) = \lim_{K \to \infty} \frac{1}{K} \,\mathrm{trace}\left(\boldsymbol{X}\right).$$

Two random matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are asymptotically free, if for all integers $n, m > 0$, and all α_{ij}, β_{ij} for which

$$\mathrm{Tr}\left(\sum_{i=0}^{m} \alpha_{1i} \boldsymbol{A}^i\right) = \cdots = \mathrm{Tr}\left(\sum_{i=1}^{m} \alpha_{ni} \boldsymbol{A}^i\right) = 0$$

and

$$\mathrm{Tr}\left(\sum_{i=0}^{m} \beta_{1i} \boldsymbol{B}^i\right) = \cdots = \mathrm{Tr}\left(\sum_{i=1}^{m} \beta_{ni} \boldsymbol{B}^i\right) = 0,$$

we have

$$\mathrm{Tr}\left(\left[\sum_{i=0}^{m} \alpha_{1i} \boldsymbol{A}^i\right]\left[\sum_{i=0}^{m} \beta_{1i} \boldsymbol{B}^i\right] \cdots \left[\sum_{i=0}^{m} \alpha_{ni} \boldsymbol{A}^i\right]\left[\sum_{i=0}^{m} \beta_{ni} \boldsymbol{B}^i\right]\right) = 0.$$

The Definition of Freeness

Two random matrices are asymptotically free, if the traces of all possible non-commutative alternating products of matrix polynomials, whose traces are zero, are zero, too.

Several random matrices are jointly asymptotically free, if the traces of all possible non-commutative alternating products of matrix polynomials, whose traces are zero, are zero, too.

It is possible and useful to define freeness of families of random matrices.

The general definition of freeness is complicated and will be given later on.

Free probability theory was invented by Dan-Virgil Voiculescu around 1986.

Dan-Virgil Voiculescu
born in Bucharest in 1949

Examples for Freeness

- Two independent Haar distributed random matrices are asymptotically free.
- Two i.i.d. Gaussian distributed random matrices are asymptotically free.
- A Haar distributed or an i.i.d. Gaussian distributed random matrix is asymptotically free from a constant matrix.

Caveat: There exist independent random matrices which are not asymptotically free and dependent random matrices which are.

Example: Let $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ be independent diagonal random matrices and let $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ be independent and Haar distributed. Then,

- $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ are **not** asymptotically free,
- **but** $\boldsymbol{H}_1\boldsymbol{D}_1\boldsymbol{H}_1^{\mathrm{H}}$ and $\boldsymbol{H}_2\boldsymbol{D}_1\boldsymbol{H}_2^{\mathrm{H}}$ **are** asymptotically free.

Independence vs. Freeness

Freeness is an asymptotic property of the eigenvectors of random matrices, but not of their eigenvalue distributions.

If the eigenvectors of two Hermitian random matrices are identical, the random matrices commute and classical probability theory is the appropriate tool to deal with them.

If the eigenvectors of two Hermitian random matrices differ, they do not commute, but they are not necessarily free.

If the product of the eigenvector matrices of two Hermitian random matrices is Haar distributed, the two random matrices are free.

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

$$\mathrm{R}(w) \stackrel{\triangle}{=} \mathrm{G}^{-1}(-w) - \frac{1}{w}$$

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.

Some R-Transforms

$$\begin{aligned}
\text{identity}: &\quad \mathrm{R}(w) = 1 \\
\text{semi-circle}: &\quad \mathrm{R}(w) = w \\
(\text{quarter-circle})^2: &\quad \mathrm{R}(w) = \frac{1}{1-w} \\
(\text{def. quarter circle})^2: &\quad \mathrm{R}(w) = \frac{\alpha}{1-w} \\
(\text{def. quarter circle})^{-2}: &\quad \mathrm{R}(w) = \frac{\alpha - 1 - \sqrt{(\alpha-1)^2 - 4w}}{2w} \\
\text{inverse semi-circle}: &\quad \mathrm{R}(w) = \frac{-1 + \sqrt{1+4w^2}}{w} \\
\text{projector}: &\quad \mathrm{R}(w) = \frac{w - 1 + \sqrt{4\alpha w + w^2 - 2w + 1}}{2w}
\end{aligned}$$

Properties of the R-Transform

The R-transform is nonlinear:

$$\mathrm{R}_{\beta \boldsymbol{X}}(w) = \beta \mathrm{R}_{\boldsymbol{X}}(\beta w)$$

The R-transform is increasing on the real line, strictly increasing unless the distribution is a single point mass.

The coefficients of the power series

$$\mathrm{R}(w) = \sum_{\ell=1}^{\infty} \beta_\ell \, w^{\ell-1}$$

are the free cumulants. The first two cumulants β_1 and β_2 are the mean and the variance of the distribution, respectively.

Effective Interference

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

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

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.

Additivity of Rank

Theorem 14 *Let $\boldsymbol{A} = \boldsymbol{A}^{\mathrm{H}}$ and $\boldsymbol{B} = \boldsymbol{B}^{\mathrm{H}}$ be free random variables. Then, the density of*

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

has a mass point at c, if and only if c can be decomposed into $c = a + b$ such that $\boldsymbol{A}$ has a mass point at a with probability α and $\boldsymbol{B}$ has a mass point at b with probability β such that

$$\alpha + \beta > 1.$$

In this case, we have

$$\Pr(c) = \alpha + \beta - 1.$$

When summing two free random matrices of normalized ranks r_1 and r_2, the normalized rank of the sum is given by $\min\{r_1 + r_2, 1\}$.

Example for Additive Free Convolution

Take two free binary random variables

$$\mathrm{p}_{\boldsymbol{A}}(x) = \mathrm{p}_{\boldsymbol{B}}(x) = \frac{1}{2}\,\delta(x-1) + \frac{1}{2}\,\delta(x+1).$$

The Stieltjes transforms of their distributions are

$$\mathrm{G}_{\boldsymbol{A}}(s) = \mathrm{G}_{\boldsymbol{B}}(s) = \frac{s}{1-s^2}.$$

Their R-transforms are given by

$$\mathrm{R}_{\boldsymbol{A}}(w) = \mathrm{R}_{\boldsymbol{B}}(w) = \frac{-1+\sqrt{1+4w^2}}{2w}.$$

Additive free convolution gives

$$\mathrm{R}_{\boldsymbol{A}+\boldsymbol{B}}(w) = \frac{-1+\sqrt{1+4w^2}}{w}.$$

which corresponds to the *inverse* semicircle law

$$\mathrm{p}_{\boldsymbol{A}+\boldsymbol{B}}(x) = \begin{cases} \frac{1}{\pi}\frac{1}{\sqrt{4-x^2}} & |x|<2 \\ 0 & \text{elsewhere} \end{cases}.$$

Example for Additive Free Convolution (cont'd)

Let n free binary RVs be summed and normalized

$$\boldsymbol{C}_n \triangleq \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \boldsymbol{A}_k.$$

We have

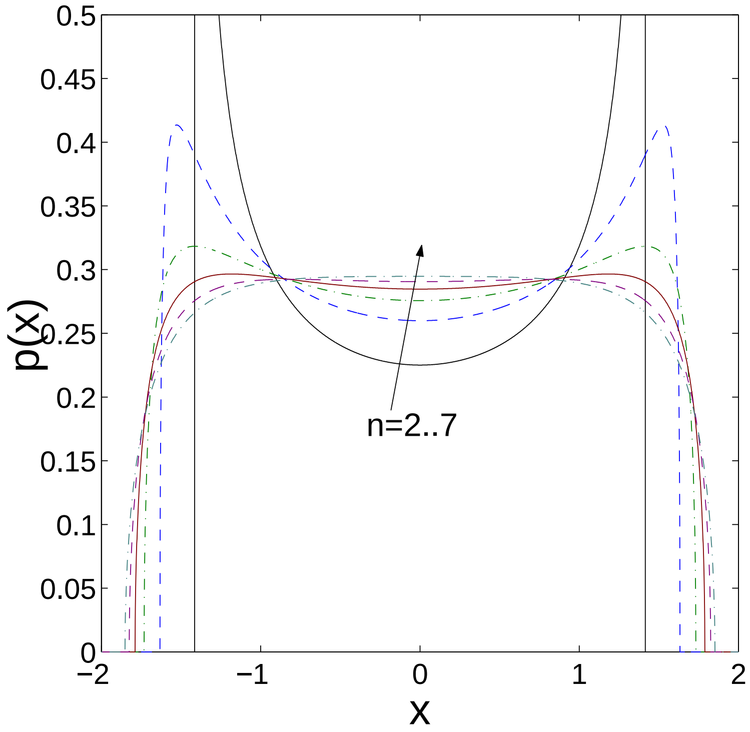
$$\mathrm{R}_{\boldsymbol{C}_n}(w) = \frac{\sqrt{n^2 + 4nw^2} - n}{2w}.$$

In the Stieltjes domain, this reads

$$\mathrm{G}_{\boldsymbol{C}_n}(s) = \frac{1}{2} \frac{(n-2)s - \sqrt{n^2 s^2 - 4n^2 + 4n}}{s^2 - n}$$

which, for $n > 1$, corresponds to the density

$$\mathrm{p}_{\boldsymbol{C}_n}(x) = \begin{cases} \frac{1}{2\pi} \frac{\sqrt{4n^2 - 4n - n^2 x^2}}{n - x^2} & |x| < 2\sqrt{1 - 1/n} \\ 0 & \text{elsewhere} \end{cases}.$$



$$\lim_{n \to \infty} \mathrm{R}_{\boldsymbol{C}_n}(w) = w$$

Free Central Limit Theorem

The R-transform is an analytic function within the neighborhood of $w = 0$. Thus,

$$\mathrm{R}_{\boldsymbol{A}_k}(w) = \sum_{\ell=0}^{\infty} \beta_{\ell,k} w^{\ell}$$

where $\beta_{0,k} = 0$ since we assume $\mathrm{Tr}(\boldsymbol{A}_k) = 0$.

$$\begin{aligned}
\lim_{n\to\infty} \mathrm{R}_{\boldsymbol{C}_n}(w) &= \lim_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \mathrm{R}_{\boldsymbol{A}_k}\left(\frac{w}{\sqrt{n}}\right) \\
&= \lim_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{\ell=1}^{\infty} \beta_{\ell,k} \left(\frac{w}{\sqrt{n}}\right)^{\ell} \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \beta_{1,k} w.
\end{aligned}$$

This is a semicircle law.

Classical Poisson Limit Theorem

The classical Poisson limit theorem states that the density of the sum of n independent random variables

$$c_n = \sum_i a_i$$

with density

$$\mathrm{p}_{a_i}(x) = \left(1 - \frac{\alpha}{n}\right)\delta(x) + \frac{\alpha}{n}\delta(x-1)$$

converges, as $n \to \infty$, to the limit

$$\mathrm{p}_{c_\infty}(x) = \mathrm{e}^{-\alpha} \sum_{k=0}^{\infty} \frac{\alpha^k}{k!}\delta(x-k)$$

which is known as the Poisson density. Similar to the central limit theorem, there is a free analog to this law.

Free Poisson Limit Theorem

Consider now a family of n free random variables $\boldsymbol{A}_i$ with density $\mathrm{p}_{\boldsymbol{A}_i}(x) = \mathrm{p}_{a_i}(x)$ and the sum

$$\boldsymbol{C}_n = \sum_i \boldsymbol{A}_i.$$

Then, as $n \to \infty$, the density of the sum converges to the squared deformed quarter circle law.

Sketch of Proof: Let $\boldsymbol{a}_i$ be an $n/\alpha \times 1$ random vector, and $\boldsymbol{A}_i = \boldsymbol{a}_i \boldsymbol{a}_i^{\mathrm{H}}$. Then, $\mathrm{p}_{a_i}(x)$ is the eigenvalue density of the matrix $\boldsymbol{A}_i$. Note that the matrix sum $\boldsymbol{C}_n = \sum_i \boldsymbol{A}_i$ is an $n/\alpha \times n/\alpha$ random covariance matrix. Thus, the singular values of $[\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n]$ follow the deformed quarter circle law for large n. Using additive free convolution, the free Poison limit theorem is shown in the R-domain, cf. effective interference.

Support of the Density Function

The boundaries of the support of the density function are the extrema of the inverse of the Stieltjes transform.

Since

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

this means

$$\frac{\partial \mathrm{R}(-w_\circ)}{\partial w_\circ} + \frac{1}{w_\circ^2} = 0$$

and $\mathrm{G}^{-1}(w_\circ)$ is a boundary point.

Maxima correspond to left boundaries of the support, minima to right boundaries.

Multiplicative Free Convolution

Let $\boldsymbol{A} = \boldsymbol{A}^{\mathrm{H}}$ and $\boldsymbol{B} = \boldsymbol{B}^{\mathrm{H}}$ be free, $\mathrm{Tr}(\boldsymbol{A}) \neq 0 \neq \mathrm{Tr}(\boldsymbol{B})$, and

$$\boldsymbol{D} = \boldsymbol{A}\boldsymbol{B} = \boldsymbol{D}^{\mathrm{H}}.$$

Then,

$$\mathrm{S}_{\boldsymbol{D}}(z) = \mathrm{S}_{\boldsymbol{A}}(z)\,\mathrm{S}_{\boldsymbol{B}}(z)\,.$$

The S-transform is defined as

$$\mathrm{S}(z) \triangleq \frac{1+z}{z}\Upsilon^{-1}(z) \qquad \text{with} \qquad \Upsilon(s) \triangleq -1 - \frac{\mathrm{G}\left(\frac{1}{s}\right)}{s}.$$

The S-transform linearizes multiplicative free convolution.

Some S-Transforms

$$
\begin{aligned}
\text{identity}: &\quad \mathrm{S}(z) = 1 \\
\text{semi-circle}: &\quad \mathrm{S}(z) = \frac{1}{\sqrt{z}} \\
(\text{quarter-circle})^2: &\quad \mathrm{S}(z) = \frac{1}{1+z} \\
(\text{def. quarter circle})^2: &\quad \mathrm{S}(z) = \frac{1}{\alpha+z} \\
(\text{def. quarter circle})^{-2}: &\quad \mathrm{S}(z) = \alpha - 1 - z \\
\text{inverse semi-circle}: &\quad \mathrm{S}(z) = \sqrt{\frac{1}{4} + \frac{1}{2z}} \\
\text{projector}: &\quad \mathrm{S}(z) = \frac{1+z}{\alpha+z}
\end{aligned}
$$

Properties of the S-Transform

The functions $z\mathrm{S}(z)$ and $z\mathrm{R}(z)$ are inverses of each other with respect to composition.

Thus, we have

$$\mathrm{R}(z) = \frac{1}{\mathrm{S}(z\mathrm{R}(z))} \qquad \Longleftrightarrow \qquad \mathrm{S}(z) = \frac{1}{\mathrm{R}(z\mathrm{S}(z))}$$

Scaling law:

$$\mathrm{S}_{\alpha \boldsymbol{X}}(z) = \alpha \mathrm{S}_{\boldsymbol{X}}(z)$$

The S-transform is decreasing on the real line, strictly decreasing unless the distribution is a single point mass.

$$\mathrm{Tr}(\boldsymbol{X}) = \frac{1}{\mathrm{S}_{\boldsymbol{X}}(0)}$$

Mass Points

Theorem 15 *Let* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *be free random variables with densities supported in* $[0;\infty)$. *Then, the density of*

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

has a mass point at $c > 0$, *if and only if* c *can be decomposed into* $c = ab$ *such that* $\boldsymbol{A}$ *has a mass point at* a *with probability* α *and* $\boldsymbol{B}$ *has a mass point at* b *with probability* β *such that*

$$\alpha + \beta > 1.$$

In this case, we have

$$\Pr(c) = \alpha + \beta - 1.$$

Theorem 16 *Let* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *be free random variables with densities supported in* $[0;\infty)$. *Let there be mass points at 0 with probabilities* α *and* β, *respectively. Then, the density of* $\boldsymbol{C} = \boldsymbol{A}\boldsymbol{B}$ *has a mass point at 0 with probability*

$$\max\{\alpha, \beta\}.$$

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 $M_{1\le n\le N}$ 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,

$$\lim_{K\to\infty} \frac{1}{K} \log \operatorname*{E}_{\boldsymbol{J}} \mathrm{e}^{K \mathrm{trace}(\boldsymbol{J}\boldsymbol{Q})} = \sum_{a=1}^{n} \int_0^{\lambda_a(\boldsymbol{Q})} \mathrm{R}_{\boldsymbol{J}}(w) \mathrm{d}w.$$

This implies

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

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

Commutative Polynomials

Let x, y be real scalar variables. The set of all n^{th} order polynomials in two commutative variables x and y is defined as

$$\mathcal{P}_n(x, y) \triangleq \left\{ \sum_{i=1}^{(n+1)^2} \beta_i x^{\ell_i} y^{m_i} : \ell_i, m_i \in \{0, 1, \ldots, n\} \wedge \beta_i \in \mathbb{R} \right\}.$$

E.g. for second order, it is canonically given by a sum with only 9 terms

$$\beta_1 x^2 y^2 + \beta_2 x^2 y + \beta_3 x y^2 + \beta_4 x^2 + \beta_5 x y + \beta_6 y^2 + \beta_7 x + \beta_8 y + \beta_9.$$

A polynomial of order n in p commutative variables can be defined by

$$\mathcal{P}_n(x_1, \ldots, x_p) \triangleq \left\{ \sum_{i=1}^{(n+1)^p} \beta_i \prod_{k=1}^{p} x_k^{\ell_{i,k}} : \ell_{i,k} \in \{0, 1, \ldots, n\} \wedge \beta_i \in \mathbb{R} \right\}.$$

Noncommutative Polynomials

Let $\boldsymbol{A}$, $\boldsymbol{B}$ be real matrices. The set of all n^{th} order polynomials in two non-commutative variables $\boldsymbol{A}$ and $\boldsymbol{B}$ is defined as

$$\mathcal{P}_n(\boldsymbol{A},\boldsymbol{B}) \triangleq \left\{ \sum_i \beta_i \prod_{k=1}^n \boldsymbol{A}^{\ell_{i,k}} \boldsymbol{B}^{m_{i,k}} : \ell_{i,k}, m_{i,k} \in \mathbb{N}_0 \wedge \sum_{k=1}^n \ell_{i,k}, \sum_{k=1}^n m_{i,k} \leq n \wedge \beta_i \in \mathbb{R} \right\}$$

E.g. for second order, it is canonically given by a sum of 19 terms

$$\beta_1 \boldsymbol{A}^2\boldsymbol{B}^2 + \beta_2 \boldsymbol{A}\boldsymbol{B}^2\boldsymbol{A} + \beta_3 \boldsymbol{A}\boldsymbol{B}\boldsymbol{A}\boldsymbol{B} + \beta_4 \boldsymbol{B}\boldsymbol{A}\boldsymbol{B}\boldsymbol{A} + \beta_5 \boldsymbol{B}\boldsymbol{A}^2\boldsymbol{B} + \beta_6 \boldsymbol{B}^2\boldsymbol{A}^2 +$$
$$+ \beta_7 \boldsymbol{A}^2\boldsymbol{B} + \beta_8 \boldsymbol{A}\boldsymbol{B}\boldsymbol{A} + \beta_9 \boldsymbol{A}\boldsymbol{B}^2 + \beta_{10} \boldsymbol{B}\boldsymbol{A}^2 + \beta_{11} \boldsymbol{B}\boldsymbol{A}\boldsymbol{B} + \beta_{12} \boldsymbol{B}^2\boldsymbol{A} +$$
$$+ \beta_{13} \boldsymbol{A}^2 + \beta_{14} \boldsymbol{A}\boldsymbol{B} + \beta_{15} \boldsymbol{B}\boldsymbol{A} + \beta_{16} \boldsymbol{B}^2 + \beta_{17} \boldsymbol{A} + \beta_{18} \boldsymbol{B} + \beta_{19} \mathbf{I}.$$

A polynomial of order n in p non-commutative variables can be defined by

$$\mathcal{P}_n(\boldsymbol{A}_1,\ldots,\boldsymbol{A}_p) \triangleq \left\{ \sum_i \beta_i \prod_{k=1}^n \prod_{q=1}^p \boldsymbol{A}_q^{\ell_{i,k,q}} : \ell_{i,k,q} \in \mathbb{N}_0 \wedge \sum_{k=1}^n \ell_{i,k,q} \leq n \wedge \beta_i \in \mathbb{R} \right\}.$$

The number of terms can be considerably large even for small values of n and p.

Freeness of Families of Random Variabes

Definition 10 *The sets* $\mathcal{Q}_1 \stackrel{\triangle}{=} \{\boldsymbol{A}_1, \dots, \boldsymbol{A}_a\}$, $\mathcal{Q}_2 \stackrel{\triangle}{=} \{\boldsymbol{B}_1, \dots, \boldsymbol{B}_b\}, \dots, \mathcal{Q}_r$ *form a free family* $(\mathcal{Q}_1, \dots, \mathcal{Q}_r)$ *if, for every sequence* $(s_1, \dots, s_k, \dots)$ *with* $s_k \in \{1, 2, \dots, r\} \, \forall k$ *and*

$$s_{k+1} \neq s_k \qquad \forall k,$$

and every sequence of polynomials $(\boldsymbol{Q}_1, \dots, \boldsymbol{Q}_k, \dots)$ *with* $\boldsymbol{Q}_k \in \mathcal{P}_\infty(\mathcal{Q}_{s_k}) \, \forall k$, *and every positive integer* n,

$$\mathrm{Tr}\,(\boldsymbol{Q}_1) = \cdots = \mathrm{Tr}\,(\boldsymbol{Q}_n) = 0 \Longrightarrow \mathrm{Tr}\,(\boldsymbol{Q}_1 \boldsymbol{Q}_2 \cdots \boldsymbol{Q}_n) = 0.$$

Note that due to the constraint on the sequences s_k adjacent factors in the product $\boldsymbol{Q}_1 \boldsymbol{Q}_2 \cdots \boldsymbol{Q}_n$ must be polynomials of different sets of the family. This reflects the non-commutative nature in the definition of freeness.

Using the Definition of Freeness

Calculate $\mathrm{Tr}(\boldsymbol{A}\boldsymbol{C}\boldsymbol{B}\boldsymbol{D})$ for the free family $\big(\{\boldsymbol{A},\boldsymbol{B}\},\{\boldsymbol{C},\boldsymbol{D}\}\big)$.

Choose the non-commutative polynomials

$$\begin{aligned}
\boldsymbol{Q}_1 &= \boldsymbol{A} - \mathrm{Tr}(\boldsymbol{A})\mathbf{I} \\
\boldsymbol{Q}_2 &= \boldsymbol{C} - \mathrm{Tr}(\boldsymbol{C})\mathbf{I} \\
\boldsymbol{Q}_3 &= \boldsymbol{B} - \mathrm{Tr}(\boldsymbol{B})\mathbf{I} \\
\boldsymbol{Q}_4 &= \boldsymbol{D} - \mathrm{Tr}(\boldsymbol{D})\mathbf{I}.
\end{aligned}$$

Polynomials with adjacent indices are built of matrices belonging to different sets of the family. Since

$$\mathrm{Tr}(\boldsymbol{Q}_k) = \mathrm{Tr}\big(\boldsymbol{X} - \mathrm{Tr}(\boldsymbol{X})\mathbf{I}\big) = 0$$

the definition of freeness implies

$$\mathrm{Tr}(\boldsymbol{Q}_1\boldsymbol{Q}_2\boldsymbol{Q}_3\boldsymbol{Q}_4) = 0.$$

Using the Definition of Freeness (cont'd)

Plug in the polynomials and solve for

$$\begin{aligned}\mathrm{Tr}(\boldsymbol{ACBD}) = {} & \mathrm{Tr}(\boldsymbol{B})\mathrm{Tr}(\boldsymbol{ACD}) + \mathrm{Tr}(\boldsymbol{D})\mathrm{Tr}(\boldsymbol{ACB}) + \mathrm{Tr}(\boldsymbol{A})\mathrm{Tr}(\boldsymbol{CBD}) + \mathrm{Tr}(\boldsymbol{C})\mathrm{Tr}(\boldsymbol{ABD}) \\ & - \mathrm{Tr}(\boldsymbol{B})\mathrm{Tr}(\boldsymbol{D})\mathrm{Tr}(\boldsymbol{AC}) - \mathrm{Tr}(\boldsymbol{A})\mathrm{Tr}(\boldsymbol{B})\mathrm{Tr}(\boldsymbol{CD}) - \mathrm{Tr}(\boldsymbol{A})\mathrm{Tr}(\boldsymbol{D})\mathrm{Tr}(\boldsymbol{CB}) \\ & - \mathrm{Tr}(\boldsymbol{C})\mathrm{Tr}(\boldsymbol{B})\mathrm{Tr}(\boldsymbol{AD}) - \mathrm{Tr}(\boldsymbol{C})\mathrm{Tr}(\boldsymbol{D})\mathrm{Tr}(\boldsymbol{AB}) - \mathrm{Tr}(\boldsymbol{A})\mathrm{Tr}(\boldsymbol{C})\mathrm{Tr}(\boldsymbol{BD}) \\ & + 3\mathrm{Tr}(\boldsymbol{A})\mathrm{Tr}(\boldsymbol{B})\mathrm{Tr}(\boldsymbol{C})\mathrm{Tr}(\boldsymbol{D}).\end{aligned}$$

We have broken down an expectation of four factors into sums of expectations of up to three factors.

The expectations of three factors can be broken down into sums of expectations of two factors. This is demonstrated at the example of $\mathrm{Tr}(\boldsymbol{ACD})$.

Using the Definition of Freeness (cont'd)

We must define the non-commutative polynomials in such a way that they belong to different sets of the free family. An appropriate definition is

$$\begin{aligned} \boldsymbol{Q}_1 &= \boldsymbol{A} - \mathrm{Tr}(\boldsymbol{A})\mathbf{I} \\ \boldsymbol{Q}_2 &= \boldsymbol{C}\boldsymbol{D} - \mathrm{Tr}(\boldsymbol{C}\boldsymbol{D})\mathbf{I}. \end{aligned}$$

Proceeding this way for all remaining matrix products involving factors belonging to different sets of the family $\left(\{\boldsymbol{A}, \boldsymbol{B}\}, \{\boldsymbol{C}, \boldsymbol{D}\}\right)$, we arrive at

$$\begin{aligned} \mathrm{Tr}(\boldsymbol{ACBD}) &= \mathrm{Tr}(\boldsymbol{A})\mathrm{Tr}(\boldsymbol{B})\mathrm{Tr}(\boldsymbol{CD}) + \mathrm{Tr}(\boldsymbol{C})\mathrm{Tr}(\boldsymbol{D})\mathrm{Tr}(\boldsymbol{AB}) \\ &\quad - \mathrm{Tr}(\boldsymbol{A})\mathrm{Tr}(\boldsymbol{B})\mathrm{Tr}(\boldsymbol{C})\mathrm{Tr}(\boldsymbol{D}). \end{aligned}$$

The key point to succeed with this procedure is to define the non-commutative polynomials in an appropriate way which simply consists of collecting all factors belonging to identical sets of the free family and subtract its expectation.

Free IID Random Matrices

Let the random matrices $\boldsymbol{H}_i, \forall i$, be square $N \times N$ with independent identically distributed entries with zero mean, variance $1/N$, and

$$\lim_{N\to\infty} \mathrm{E}\left|\sqrt{N}\,(\boldsymbol{H}_i)_{11}\right|^{2m} < \infty \quad \forall m > 0$$

Moreover, let $\boldsymbol{X}_j, \forall j$, be $N \times N$ matrices with upper bounded norm and a limit distribution as $N \to \infty$.

Then the family

$$\left(\left\{ \boldsymbol{X}_1, \boldsymbol{X}_1^{\mathrm{H}}, \boldsymbol{X}_2, \boldsymbol{X}_2^{\mathrm{H}}, \ldots \right\}, \left\{ \boldsymbol{H}_1, \boldsymbol{H}_1^{\mathrm{H}} \right\}, \left\{ \boldsymbol{H}_2, \boldsymbol{H}_2^{\mathrm{H}} \right\}, \ldots \right)$$

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

Free Hermitian Random Matrices

Let the random matrices $\boldsymbol{H}_i, \forall i$, be $N \times K$ with independent identically distributed entries with zero mean, variance $1/N$, and

$$\lim_{N\to\infty} \mathrm{E}\left|\sqrt{N}\left(\boldsymbol{H}_i\right)_{11}\right|^{2m} < \infty \quad \forall m > 0$$

and let the matrices $\boldsymbol{Y}_i, \forall i$, be $K \times K$, Hermitian and independent of $\boldsymbol{H}_i, \forall i$. Let

$$\boldsymbol{S}_i = \boldsymbol{H}_i \boldsymbol{Y}_i \boldsymbol{H}_i^{\mathrm{H}}, \forall i.$$

Moreover, let $\boldsymbol{X}_j, \forall j$, be $N \times N$ matrices with upper bounded norm and a limit distribution as $N \to \infty$.

Then the family

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

is almost surely asymptotically free as $N, K \to \infty$ with $\alpha = K/N$ fixed.

Free Unitary Random Matrices

Let the random matrices $\boldsymbol{T}_i, \forall i$, be $N \times N$ Haar distributed random matrices.

Moreover, let $\boldsymbol{X}_j, \forall j$, be an $N \times N$ matrices with upper bounded norm and a limit distribution as $N \to \infty$.

Then, the family

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

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

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

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

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 $N \to \infty$.

R-Diagonal Random Matrices

Definition 11 *A random variable* $\boldsymbol{X}$ *is called R-diagonal, if it can be decomposed as*

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{Y}$$

where $\boldsymbol{Y} = \sqrt{\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}}$ *and* $\boldsymbol{U}$ *is Haar distributed and free of* $\boldsymbol{Y}$.

Lemma 3 *Asymptotically large bi-unitarily invariant matrices are R-diagonal.*

Theorem 18 *The distribution of R-diagonal random matrices is circularly symmetric in the complex plane.*

R-Diagonal Additive Free Convolution

Theorem 19 *Let the asymptotically free random matrices* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *be R-diagonal and denote the respective asymptotic singular value distributions by* $\mathrm{P}_{\boldsymbol{A}}(x)$ *and* $\mathrm{P}_{\boldsymbol{B}}(x)$*. Define the symmetrization of a density by*

$$\tilde{\mathrm{p}}(x) = \frac{\mathrm{p}(x) + \mathrm{p}(-x)}{2}.$$

Then, $\boldsymbol{A} + \boldsymbol{B}$ *is R-diagonal and we have*

$$\tilde{\mathrm{R}}_{\boldsymbol{A}+\boldsymbol{B}}(w) = \tilde{\mathrm{R}}_{\boldsymbol{A}}(w) + \tilde{\mathrm{R}}_{\boldsymbol{B}}(w)$$

with $\tilde{\mathrm{R}}(w)$ *denoting the R-transform of* $\tilde{\mathrm{P}}(x)$*.*

This is an addition law for bi-unitarily invariant random matrices.

Circularly Symmetric Distributions

Theorem 20 *Let the random variable* $\boldsymbol{H}$ *be R-diagonal. Let* $\mathrm{S}_{\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}}(s)$ *denote the S-transform of the distribution of* $\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}$ *and define the function*

$$f(s) = \frac{1}{\sqrt{\mathrm{S}_{\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}}}(s-1)}}.$$

Then, the distribution of $\boldsymbol{H}$ *is circularly symmetric and given by*

$$\mathrm{p}_{\boldsymbol{H}}(z) = \frac{1}{2\pi z f'\left[f^{-1}(z)\right]}$$

with $f'(s) = \mathrm{d}f(s)/\mathrm{d}s$ *wherever the density is positive and continuous.*

The S-transform of $\boldsymbol{H}\boldsymbol{H}$ can be found from the complex-valued distribution $\mathrm{p}_{\boldsymbol{H}}(z)$ solving a differential equation.

Haagerup-Larsen Law

Theorem 21 *Let the random variables $\boldsymbol{A}_n$ have the same distribution, be R-diagonal, and free of each other for all n. Then, the distributions of*

$$\prod_{n=1}^{N} \boldsymbol{A}_n$$

and $\boldsymbol{A}_1^N$ are identical.

Free identically distributed R-diagonal random variables behave with respect to multiplication as if they were identical.

Chapter 6:

The Replica Method

The Return of Physics

Since Boltzmann, physicists have studied the behavior of systems with interactions of many particles.

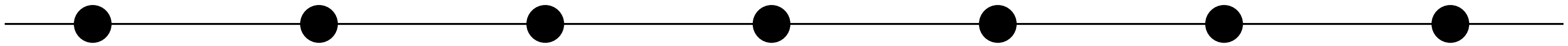
Particular correspondences can be drawn between spin glasses (amorph magnetic materials, e.g. the magnetic surface of a hard disk drive) and communication systems. The binary nature of the bits corresponds to the quantum-mechanical constraints of electron spins to $\pm\frac{1}{2}$.

Spin glass theory is both very rich and very complicated. Various methods have been proposed by physicists to analyze them:

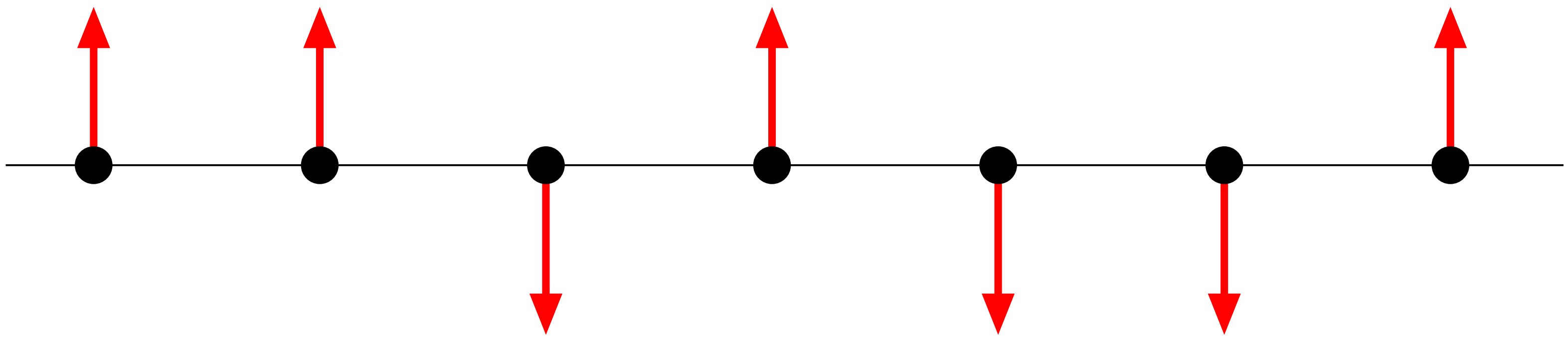
- The replica method
- The cavity method
- ...
- The TAP approach
- Gauge Theory

This course will be restricted in scope to the replica method.

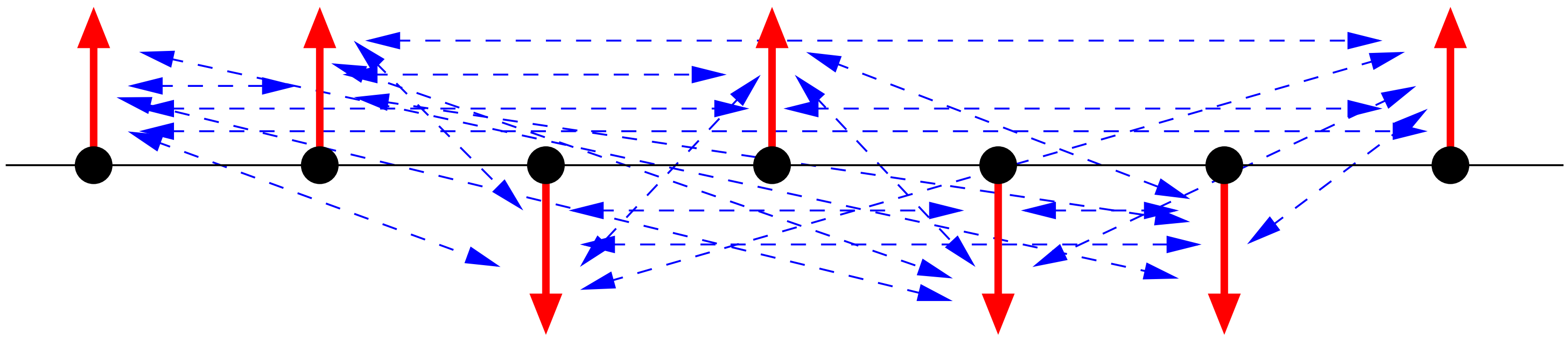
Spin Glasses



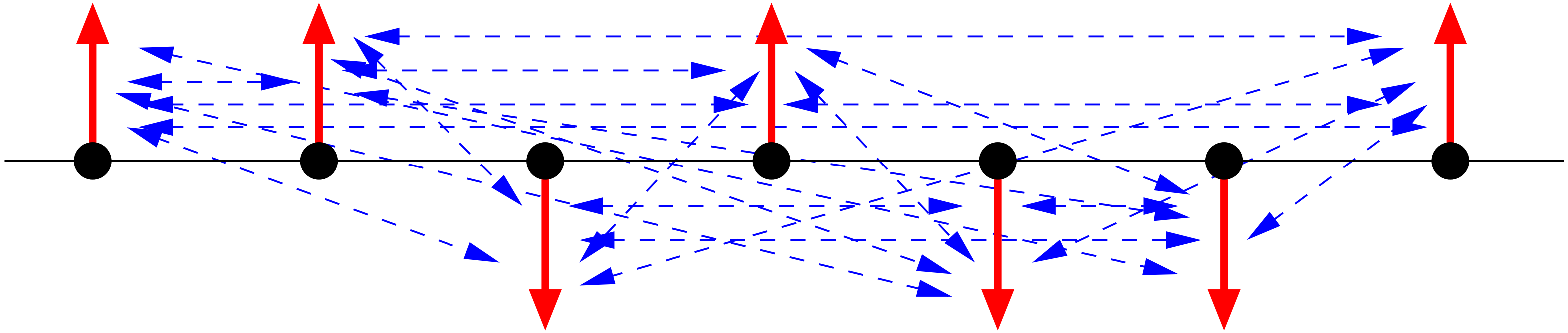
Spin Glasses



Spin Glasses



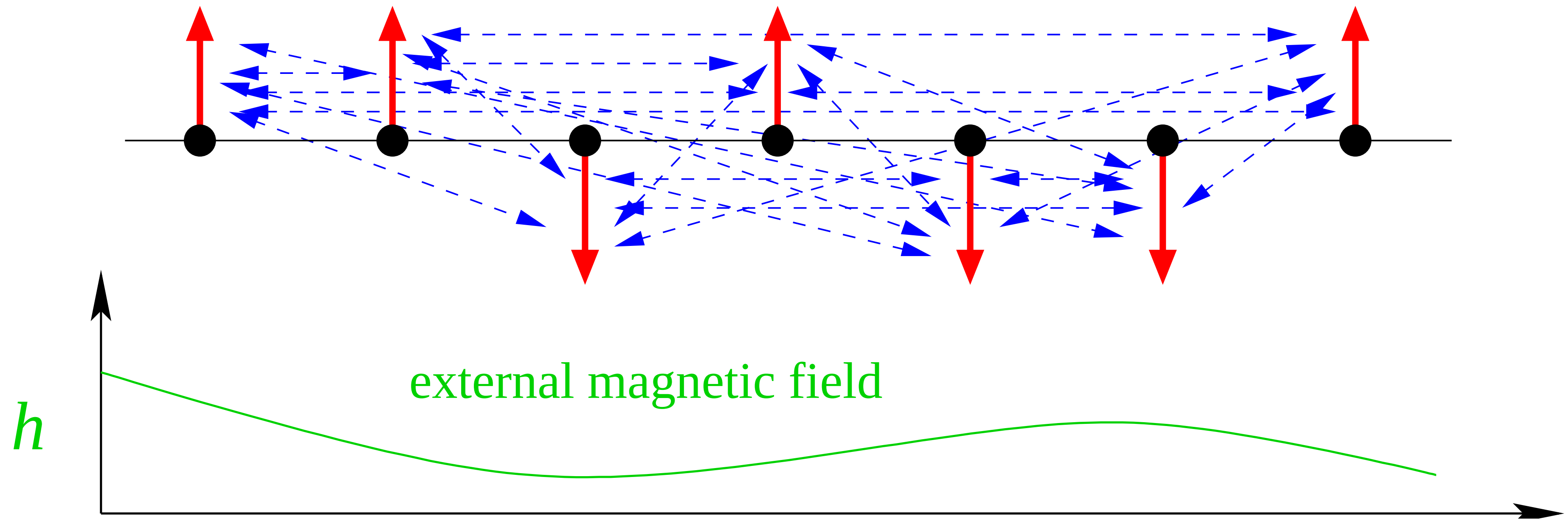
Spin Glasses



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:

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

Optimal Detection of Vector Channel

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

Best estimate for transmitted data:

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

with

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

Optimal Detection of Vector Channel

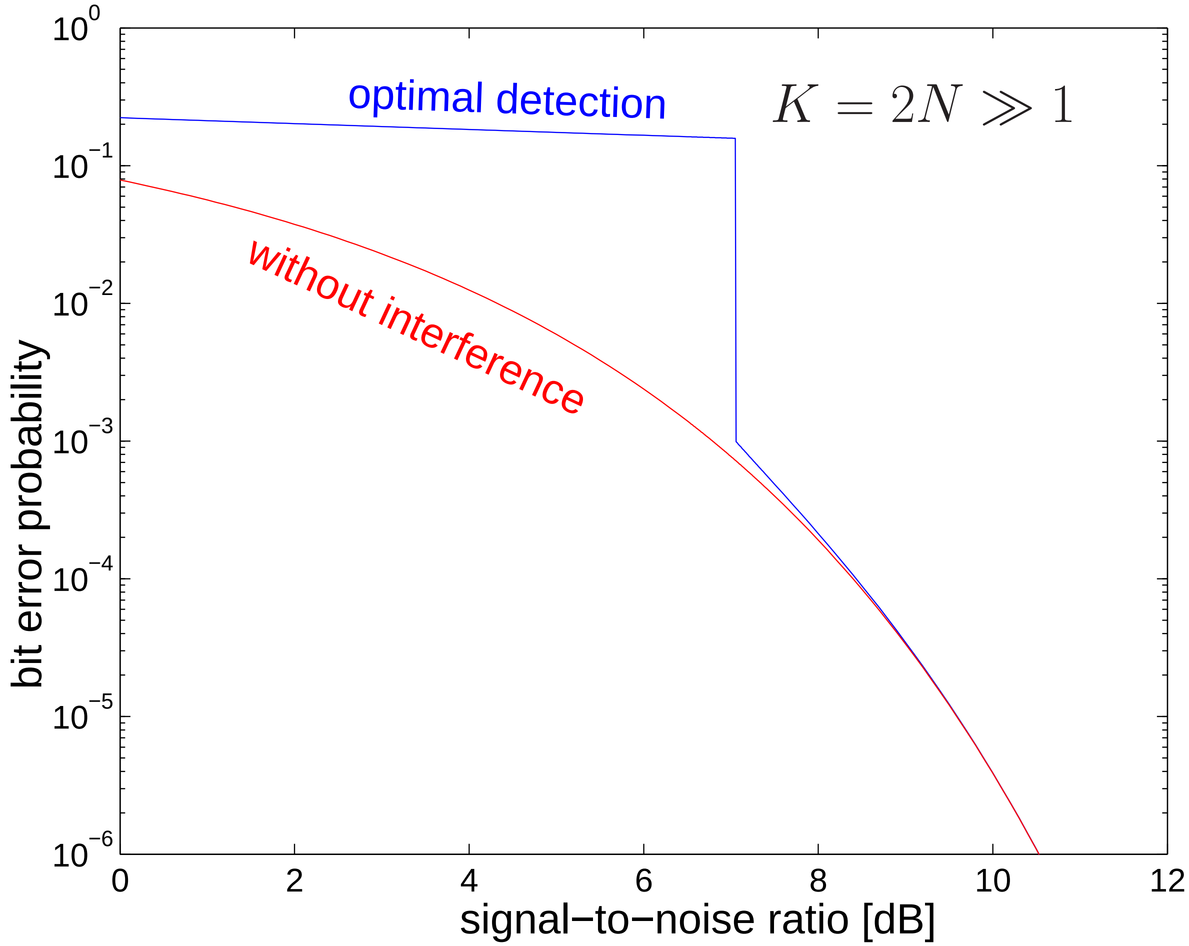
$$\boldsymbol{y} = \boldsymbol{S}\boldsymbol{x} + \boldsymbol{n}$$

Best estimate for transmitted data:

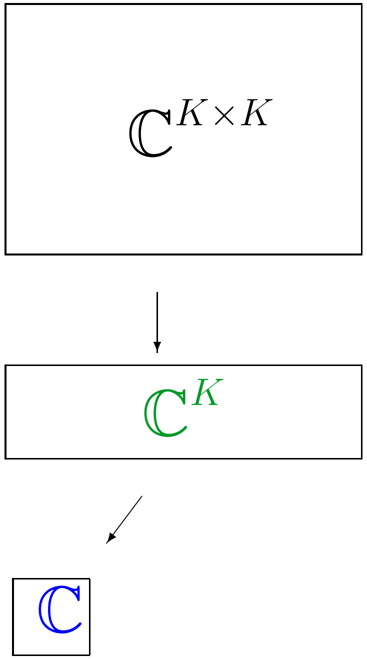
$$\begin{aligned}
\hat{\boldsymbol{x}} &= \operatorname*{argmin}_{\boldsymbol{x}\in\{\pm1\}^K} ||\boldsymbol{y} - \boldsymbol{S}\boldsymbol{x}|| \\
&= \operatorname*{argmin}_{\boldsymbol{x}\in\{\pm1\}^K} -\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{x} - \boldsymbol{h}^{\mathrm{T}}\boldsymbol{x} \qquad \text{with} \qquad \begin{aligned} \boldsymbol{R} &= -2\boldsymbol{S}^{\mathrm{T}}\boldsymbol{S} \\ \boldsymbol{h} &= 2\boldsymbol{S}^{\mathrm{T}}\boldsymbol{y} \end{aligned} \\
&= \operatorname*{argmin}_{\boldsymbol{x}\in\{\pm1\}^K} -\sum_i \sum_{j<i} r_{ij} x_i x_j - \sum_i h_i x_i
\end{aligned}$$

Minimization of the energy function of a spin glass!

A Phase Transition in Random CDMA



Large Systems



$\mathcal{O}\left(K^2\right)$ interactions

$\uparrow$

$\mathcal{O}(K)$ microscopic objects

$\downarrow$

$\mathcal{O}(1)$ macroscopic variables

Macroscopic variables are self-averaging.

Boltzmann Distribution

The Thermodynamic Equilibrium maximizes the entropy

$$H(X) = -\sum_i \Pr(x_i) \log \Pr(x_i)$$

for given constant energy

$$E(X) = \sum_i ||x_i|| \Pr(x_i)$$

yielding the *Boltzmann distribution*

$$\Pr(x_i) = \frac{e^{-\frac{1}{T}||x_i||}}{\sum_i e^{-\frac{1}{T}||x_i||}}.$$

Ludwig Eduard Boltzmann
born in Vienna in 1844

Free Energy

Since the energy is constant, we can minimize the free energy

$$F(X) \stackrel{\triangle}{=} E(X) - TH(X)$$

instead of maximizing entropy. This is often less complicated.

With the Boltzmann distribution, the free energy is given by

$$F(X) = -T \log \left[\sum_i \mathrm{e}^{-\frac{1}{T}||x_i||} \right].$$

It depends only on the partition function.

The free energy is self-averaging.

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_ks be uniformly bounded from above and the U_ks 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*

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

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

Average Free Energy

When analyzing a random system, we evaluate the average free energy

as Shannon analyzed the average performance of all codes.

Free Energy for a Random Parameter

Consider a self-averaging random parameter, e.g. a spreading matrix.

$$
\begin{aligned}
F(X|y_j) &= \mathop{\mathrm{E}}_{Y} F(X|Y) \\
&= -T \mathop{\mathrm{E}}_{Y} \log \left[\sum_i \mathrm{e}^{-\frac{1}{T}||x_i||} \right]
\end{aligned}
$$

The energy function depends on the random parameter y_j.

The expectation of a logarithm is a hard problem.

Replica Continuity

$$\mathop{\mathrm{E}}_{Y} \log(Y) = \lim_{n\to 0} \frac{\partial}{\partial n} \mathop{\mathrm{E}}_{Y} Y^n$$

Evalute the n^{th} moments for integer n and assume analytic continuity for the limit.

More general, we have

$$\mathop{\mathrm{E}}_{Y} \log \int_{\mathbb{R}} f(x, Y) dx = \lim_{n\to 0} \frac{\partial}{\partial n} \mathop{\mathrm{E}}_{Y} \left[\int_{\mathbb{R}} f(x, Y) dx \right]^n$$

Mark Kac introduced the replica method at the Theoretical Physics Seminar in Trondheim in 1968.

Mark Kac
born in 1914 in Krzemieniec

Replica Continuity (cont'd)

$$\operatorname*{E}_{Y} \log \int_{\mathbb{R}} f(x, Y) dx = \lim_{n \to 0} \frac{\partial}{\partial n} \operatorname*{E}_{Y} \left[\int_{\mathbb{R}} f(x, Y) dx \right]^n$$

With

$$\left(\int_{\mathbb{R}} g(x) dx \right)^n = \prod_{a=1}^{n} \int_{\mathbb{R}} g(x_a) dx_a$$

we finally get

$$\operatorname*{E}_{Y} \log \int_{\mathbb{R}} f(x, Y) dx = \lim_{n \to 0} \frac{\partial}{\partial n} \operatorname*{E}_{Y} \prod_{a=1}^{n} \int_{\mathbb{R}} f(x_a, Y) dx_a$$

Replica Symmetry

Throughout the calculations, we solve integrals of the form

$$I = \frac{1}{K} \log \int_{\mathbb{R}^2} \mathrm{e}^{K f(x_1,x_2)} dx_1 dx_2 \rightarrow \max_{x_1,x_2} f(x_1, x_2)$$

for $K \rightarrow \infty$ by *saddle point integration*.

If the maximization is too tedious, we assume *replica symmetry*:

$$\max_{x_1,x_2} f(x_1, x_2) = \max_{x} f(x, x)$$

Replica symmetry is a strong assumption and not always valid.

Hubbard-Stratonovich Transform

When dealing with quadratic energy functions, the problem of integrating over multivariate Gauß (Gauss) kernels occurs.

The multivariate Gauß kernel can be reduced to a univariate one by means of the Hubbard-Stratonovich transform

$$\exp\left(\frac{y}{2}\right) = \int \exp\left(\pm\sqrt{y}z\right) \mathrm{D}z \qquad \forall y \in \mathbb{R}$$

with $\mathrm{D}z$ denoting the Gaussian measure

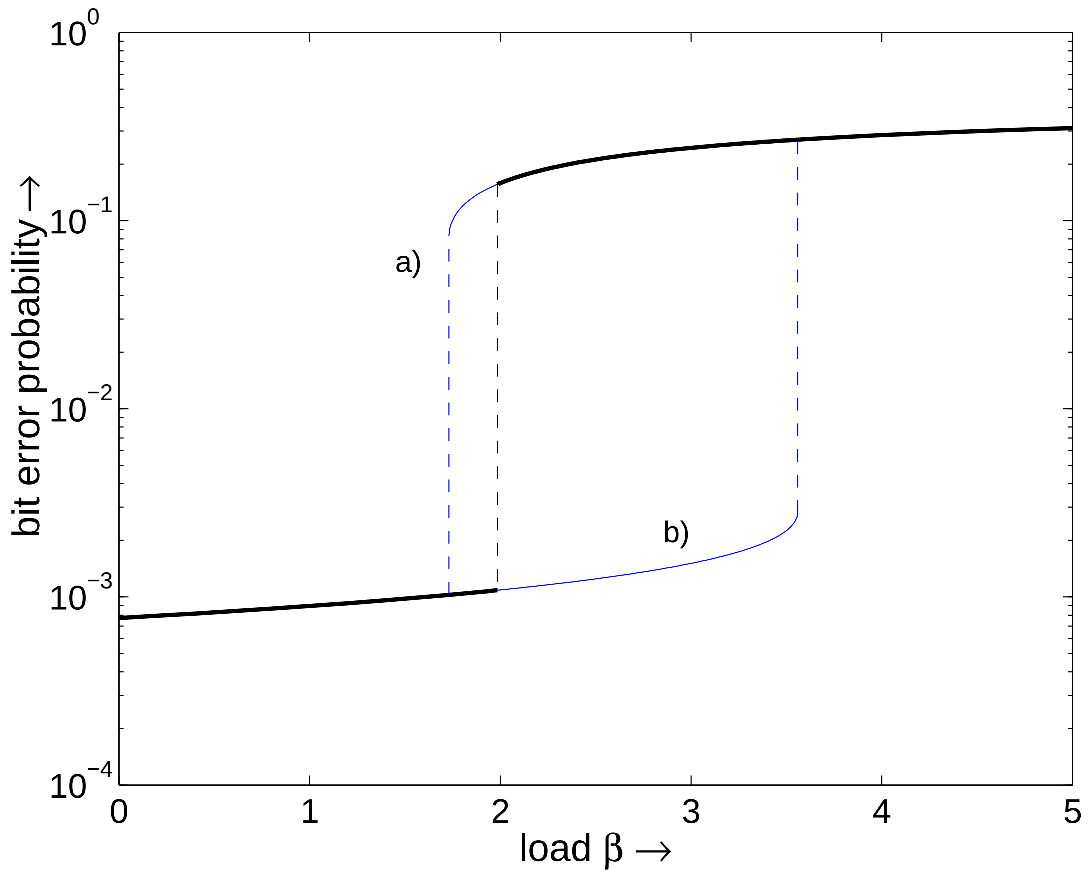
$$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \mathrm{d}z$$

In replica calculations, y is the square of a sum of replicated variables.

Phase Transitions

If the final equations allow for multiple solutions, the correct solution is identified by minimizing the free energy.

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

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

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

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

is smallest.

Chapter 7:

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. It includes multi-carrier CDMA transmission with users of arbitrary powers in frequency-selective fading as special case.

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

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

Moreover, let the detector be characterized by the assumed conditional probability distribution

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

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

Let the entries of $\boldsymbol{H}$ be independent zero-mean with vanishing odd order moments and variances w_{ck}^2/N for row c and column k.

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

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

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

With (1) and (2), the definition of free energy, and replica continuity, we find for the free energy per user

$$\begin{aligned}
\frac{\mathrm{F}(\boldsymbol{x})}{K} &= -\frac{1}{K}\mathop{\mathrm{E}}_{\boldsymbol{H}}\int\int_{\mathbb{R}^N}\log\left(\int \mathrm{e}^{-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})}\mathrm{d}\breve{\mathrm{P}}_{\boldsymbol{x}}(\boldsymbol{x})\right)\frac{\mathrm{e}^{-\frac{1}{2\sigma_0^2}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})}}{(2\pi\sigma_0^2)^{\frac{N}{2}}}\mathrm{d}\boldsymbol{y}\mathrm{d}\mathrm{P}_{\boldsymbol{x}}(\boldsymbol{x}) \\
&= -\frac{1}{K}\lim_{n\to 0}\frac{\partial}{\partial n}\log\mathop{\mathrm{E}}_{\boldsymbol{H}}\int\int_{\mathbb{R}^N}\left(\int \mathrm{e}^{-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})}\mathrm{d}\breve{\mathrm{P}}_{\boldsymbol{x}}(\boldsymbol{x})\right)^n \\
&\qquad \frac{\mathrm{e}^{-\frac{1}{2\sigma_0^2}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x})}}{(2\pi\sigma_0^2)^{\frac{N}{2}}}\mathrm{d}\boldsymbol{y}\mathrm{d}\mathrm{P}_{\boldsymbol{x}}(\boldsymbol{x}) \\
&= -\frac{1}{K}\lim_{n\to 0}\frac{\partial}{\partial n}\log\underbrace{\int\frac{\int_{\mathbb{R}^N}\mathop{\mathrm{E}}_{\boldsymbol{H}}\prod_{a=0}^{n}\mathrm{e}^{-\frac{1}{2\sigma_a^2}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}_a)^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}_a)}\mathrm{d}\boldsymbol{y}}{(2\pi\sigma_0^2)^{\frac{N}{2}}}\prod_{a=0}^{n}\mathrm{d}\mathrm{P}_a(\boldsymbol{x}_a)}_{\triangleq\,\Xi_n} \qquad (3)
\end{aligned}$$

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

The integral in (3) is given by

$$\Xi_n = \int \prod_{c=1}^{N} \frac{\int\limits_{\mathbb{R}} \mathop{\mathrm{E}}_{\boldsymbol{H}} \prod_{a=0}^{n} \exp\left[-\frac{1}{2\sigma_a^2} \left(y_c - \sum_{k=1}^{K} h_{ck} x_{ak} \right)^2 \right] \mathrm{d}y_c}{\sqrt{2\pi}\sigma_0} \prod_{a=0}^{n} \mathrm{dP}_a(\boldsymbol{x}_a)\,, \qquad (4)$$

with y_c, x_{ak}, and h_{ck} denoting the c^{th} component of $\boldsymbol{y}$, the k^{th} component of $\boldsymbol{x}_a$, and the $(c,k)^{\text{th}}$ entry of $\boldsymbol{H}$, respectively. The integrand depends on $\boldsymbol{x}_a$ only through

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

These quantities v_{ac} can be regarded, in the limit $K \to \infty$ as jointly Gaussian random variables with zero mean and covariances

$$Q_{ab}[c] = \mathop{\mathrm{E}}_{\boldsymbol{H}} v_{ac} v_{bc} = \frac{1}{K} \boldsymbol{x}_a \overset{(c)}{\bullet} \boldsymbol{x}_b \qquad (5)$$

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

In order to perform the integration in (4), the $K(n+1)$-dimensional space spanned by the replicas and the vector $\boldsymbol{x}_0$ is split into subshells

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

where the inner product of two different vectors $\boldsymbol{x}_a$ and $\boldsymbol{x}_b$ is constant.[1]

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 $K \to \infty$ [2]

$$\Xi_n = \int\limits_{\mathbb{R}^{N(n+1)(n+2)/2}} \mathrm{e}^{K\mathcal{I}\{Q[\cdot]\}} \prod_{c=1}^{N} \mathrm{e}^{\mathcal{G}\{Q[c]\}} \prod_{a \le b} \mathrm{d}Q_{ab}[c], \tag{6}$$

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

[1] The notation $f\{Q[\cdot]\}$ expresses the dependency of the function $f(\cdot)$ on all $Q_{ab}[c], 0 \le a \le b \le n, 1 \le c \le N$.
[2] The notation $\prod_{a \le b}$ is used as shortcut for $\prod_{a=0}^{n} \prod_{b=a}^{n}$.

In (6),

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

denotes the probability weight of the subshell and

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

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 variables $Q_{ab}[\cdot]$. In the following the blue and green terms in (6) are evaluated separately.

First, we calculate the measure $\mathrm{e}^{K\mathcal{I}\{Q[\cdot]\}}$. We write the Dirac measure

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

as the inverse Laplace transform of a constant with $\mathcal{J} = (t - \mathrm{j}\infty; t + \mathrm{j}\infty)$.
Then, the measure $\mathrm{e}^{K\mathcal{I}\{Q[\cdot]\}}$ can be expressed as

$$\begin{aligned} \mathrm{e}^{K\mathcal{I}\{Q[\cdot]\}} &= \int \left[\prod_{c=1}^{N} \prod_{a \le b} \int_{\mathcal{J}} \mathrm{e}^{\tilde{Q}_{ab}[c]\left(\frac{\boldsymbol{x}_a \overset{(c)}{\bullet} \boldsymbol{x}_b}{N} - \alpha Q_{ab}[c]\right)} \frac{\mathrm{d}\tilde{Q}_{ab}[c]}{2\pi \mathrm{j}}\right] \prod_{a=0}^{n} \mathrm{dP}_a(\boldsymbol{x}_a) \\ &= \int_{\mathcal{J}^{N(n+2)(n+1)/2}} \mathrm{e}^{-\alpha \sum_{c=1}^{N} \sum_{a \le b} \tilde{Q}_{ab}[c] Q_{ab}[c]} \left(\prod_{k=1}^{K} M_k\left\{\tilde{Q}[\cdot]\right\}\right) \prod_{c=1}^{N} \prod_{a \le b} \frac{\mathrm{d}\tilde{Q}_{ab}[c]}{2\pi \mathrm{j}} \qquad (7) \end{aligned}$$

with

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

In the limit of $K \to \infty$ one of the exponential terms in (6) will dominate over all others. Only the maximum value of the correlation $Q_{ab}[c]$ is relevant for calculation of the integral.

We assume that the replicas within the dominant subshell are symmetric (replica symmetry). Thus, the maximum values of the correlations $Q_{ab}[c]$ are identical for all positive $a \neq b$. The same applies to the the correlations $Q_{a0}[c]$.

Hereby, we reduce the number of different correlation variables from $(n+1)(n+2)/2$ to four per chip time and set $Q_{00}[c] = p_{0c}$, $Q_{0a}[c] = m_c, \forall a \neq 0$, $Q_{aa}[c] = p_c, \forall a \neq 0$, $Q_{ab}[c] = q_c, \forall 0 \neq a \neq b \neq 0$.

We apply the same idea to the correlation variables in the Laplace domain and set $\tilde{Q}_{00}[c] = G_{0c}/2$, $\tilde{Q}_{aa}[c] = G_c/2, \forall a \neq 0$, $\tilde{Q}_{0a}[c] = E_c, \forall a \neq 0$, and $\tilde{Q}_{ab}[c] = F_c, \forall 0 \neq a \neq b \neq 0$.

At this point the crucial benefit of the replica method becomes obvious. Assuming replica continuity, we have managed to reduce the evaluation of a continuous function to sampling it at integer points. Assuming replica symmetry we have reduced the task of evaluating infinitely many integer points to calculating 8 different correlations (4 of them in the original and 4 of them in the Laplace domain).

The assumption of replica symmetry leads to

$$\sum_{a\le b} \tilde{Q}_{ab}[c] Q_{ab}[c] = nE_c m_c + \frac{n(n-1)}{2} F_c q_c + \frac{G_{0c} p_{0c}}{2} + \frac{n}{2} G_c p_c \tag{8}$$

and

$$M_k\{E,F,G,G_0\} = \int\limits_{\mathbb{R}^{n+1}} \mathrm{e}^{\frac{1}{N}\sum\limits_{c=1}^{N} w_{ck}^2 \left(\frac{G_{0c}}{2} x_{0k}^2 + \sum\limits_{a=1}^{n} E_c x_{0k} x_{ak} + \frac{G_c}{2} x_{ak}^2 + \sum\limits_{b=a+1}^{n} F_c x_{ak} x_{bk} \right)} \prod_{a=0}^{n} \mathrm{dP}_a(x_{ak})$$

$$= \int\limits_{\mathbb{R}^{n+1}} \mathrm{e}^{\frac{\tilde{G}_{0k}}{2} x_{0k}^2 + \sum\limits_{a=1}^{n} \tilde{E}_k x_{0k} x_{ak} + \frac{\tilde{G}_k}{2} x_{ak}^2 + \sum\limits_{b=a+1}^{n} \tilde{F}_k x_{ak} x_{bk}} \prod_{a=0}^{n} \mathrm{dP}_a(x_{ak}) \tag{9}$$

where

$$\tilde{E}_k \triangleq \frac{1}{N}\sum_{c=1}^{N} E_c w_{ck}^2 \ , \qquad \tilde{F}_k \triangleq \frac{1}{N}\sum_{c=1}^{N} F_c w_{ck}^2 \tag{10}$$

$$\tilde{G}_k \triangleq \frac{1}{N}\sum_{c=1}^{N} G_c w_{ck}^2 \ , \qquad \tilde{G}_{0k} \triangleq \frac{1}{N}\sum_{c=1}^{N} G_{0c} w_{ck}^2 . \tag{11}$$

Next, we write $M_k\{E,F,G,G_0\}$ with a single quadratic term in the exponential argument by completing the square. Then, we apply the Hubbard-Stratonovich transform to linearize the exponential argument.

$$
\begin{aligned}
M_k\{E,F,G,G_0\} &= \\
&= \int \mathrm{e}^{\frac{\tilde G_{0k}}{2}x_{0k}^2+\sum\limits_{a=1}^{n}\tilde E_k x_{0k}x_{ak}+\frac{\tilde G_k}{2}x_{ak}^2+\sum\limits_{b=a+1}^{n}\tilde F_k x_{ak}x_{bk}} \prod_{a=0}^{n}\mathrm{dP}_a(x_{ak}) \\
&= \int \mathrm{e}^{\frac{\tilde G_{0k}}{2}x_{0k}^2+\frac{\tilde F_k}{2}\left(\sum\limits_{a=1}^{n}x_{ak}\right)^2+\sum\limits_{a=1}^{n}\tilde E_k x_{0k}x_{ak}+\frac{\tilde G_k-\tilde F_k}{2}x_{ak}^2} \prod_{a=0}^{n}\mathrm{dP}_a(x_{ak}) \\
&= \iint \mathrm{e}^{\frac{\tilde G_{0k}}{2}x_{0k}^2+\sum\limits_{a=1}^{n}\tilde E_k x_{0k}x_{ak}+\sqrt{\tilde F_k}zx_{ak}+\frac{\tilde G_k-\tilde F_k}{2}x_{ak}^2}\mathrm{D}z \prod_{a=0}^{n}\mathrm{dP}_a(x_{ak}) \\
&= \int \mathrm{e}^{\frac{\tilde G_{0k}}{2}x_k^2}\int\!\left(\int \mathrm{e}^{\tilde E_k x_k \breve x_k+\sqrt{\tilde F_k}z\breve x_k+\frac{\tilde G_k-\tilde F_k}{2}\breve x_k^2}\mathrm{d}\breve{\mathrm{P}}_{\breve x_k}(\breve x_k)\right)^n \mathrm{D}z\,\mathrm{dP}_{x_k}(x_k)
\end{aligned}
\tag{12}
$$

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

Second, we evaluate $\mathrm{e}^{\mathcal{G}\{Q[c]\}}$ in (6). We use the replica symmetry to construct the correlated Gaussian random variables v_{ac} out of independent zero-mean, unit-variance Gaussian random variables u_c, t_c, z_{ac} by

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

With that substitution, we get

$$\begin{aligned} \mathrm{e}^{\mathcal{G}(m_c,q_c,p_c,p_{0c})} &= \frac{1}{\sqrt{2\pi}\sigma_0}\int\limits_{\mathbb{R}^2}\int\limits_{\mathbb{R}} \exp\left[-\frac{\alpha}{2\sigma_0^2}\left(u_c\sqrt{p_{0c} - \frac{m_c^2}{q_c}} - \frac{t_c m_c}{\sqrt{q_c}} - \frac{y_c}{\sqrt{\alpha}}\right)^2\right] \mathrm{D}u_c \\ &\quad \times \left[\int\limits_{\mathbb{R}} \exp\left[-\frac{\alpha}{2\sigma^2}\left(z_c\sqrt{p_c - q_c} - t_c\sqrt{q_c} - \frac{y_c}{\sqrt{\alpha}}\right)^2\right] \mathrm{D}z_c\right]^n \mathrm{D}t_c\,\mathrm{d}y_c \\ &= \sqrt{\frac{(1+\frac{\alpha}{\sigma^2}(p_c - q_c))^{1-n}}{1+\frac{\alpha}{\sigma^2}(p_c-q_c) + n\frac{\alpha}{\sigma^2}\left(\frac{\sigma_0^2}{\alpha} + p_{0c} - 2m_c + q_c\right)}} \qquad (13) \end{aligned}$$

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

Since the integral in (6) is dominated by the maximum argument of the exponential function, the derivatives of

$$\frac{1}{N}\sum_{c=1}^{N}\left(\mathcal{G}\{Q[c]\}-\alpha\sum_{a\le b}\tilde{Q}_{ab}[c]Q_{ab}[c]\right) \tag{14}$$

with respect to m_c, q_c, p_c and p_{0c} must vanish as $N \to \infty$. Taking derivatives after plugging (8) and (13) into (14), solving for E_c, F_c, G_c, and G_{0c} and letting $n \to 0$ yields for all c

$$E_c = \frac{1}{\sigma^2+\alpha(p_c-q_c)} \tag{15}$$

$$F_c = \frac{\sigma_0^2+\alpha\left(p_{0c}-2m_c+q_c\right)}{[\sigma^2+\alpha(p_c-q_c)]^2}$$

$$G_c = F_c - E_c \tag{16}$$

$$G_{0c} = 0. \tag{17}$$

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

Gaussian Prior Distribution

Assume the Gaussian prior

$$\mathrm{p}_a(x_{ak}) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x_{ak}^2/2} \qquad \forall a.$$

The integration in (12) can be performed explicitly and we find

$$M_k\left(\tilde{E}_k, \tilde{F}_k, \tilde{G}_k, \tilde{G}_{0k}\right) = \sqrt{\frac{\left(1+\tilde{F}_k-\tilde{G}_k\right)^{1-n}}{\left(1-\tilde{G}_{0k}\right)\left(1+\tilde{F}_k-\tilde{G}_k-n\tilde{F}_k\right)-n\tilde{E}_k^2}}. \tag{18}$$

In the large system limit, the integral in (7) is dominated by that value of the integration variable which maximizes the argument of the exponential function. Thus, partial derivations of

$$\log \prod_{k=1}^{K} M_k\left(\tilde{E}_k, \tilde{F}_k, \tilde{G}_k, \tilde{G}_{0k}\right) - \alpha \sum_{c=1}^{N} nE_c m_c + \frac{n(n-1)}{2} F_c q_c + \frac{G_{0c} p_{0c}}{2} + \frac{n}{2} G_c p_c \tag{19}$$

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 yields

$$m_c = \frac{1}{K}\sum_{k=1}^{K} w_{ck}^2 \frac{\tilde{E}_k}{1+\tilde{E}_k} \tag{20}$$

$$q_c = \frac{1}{K}\sum_{k=1}^{K} w_{ck}^2 \frac{\tilde{E}_k^2+\tilde{F}_k}{\left(1+\tilde{E}_k\right)^2} \tag{21}$$

$$p_c = \frac{1}{K}\sum_{k=1}^{K} w_{ck}^2 \frac{\tilde{E}_k^2+\tilde{E}_k+\tilde{F}_k+1}{\left(1+\tilde{E}_k\right)^2} \tag{22}$$

$$p_{0c} = \frac{1}{K}\sum_{k=1}^{K} w_{ck}^2 \tag{23}$$

in the limit $n \to 0$ with (16) and (17).

Surprisingly, if we let the true prior to be binary and only the replicas to be Gaussian we also find (20) to (23). Note from Chapter 1 that this setting corresponds to linear MMSE detection.

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

$$
\begin{aligned}
-\frac{1}{K}\frac{\partial}{\partial n}\log \Xi_n &= \frac{1}{K}\frac{\partial}{\partial n}\sum_{c=1}^{N}\left[-\mathcal{G}(m_c,q_c,p_c,p_{0c}) + \alpha n E_c m_c + \frac{\alpha n(n-1)}{2}F_c q_c + \frac{\alpha n}{2}G_c p_c\right] \\
&\quad -\sum_{k=1}^{K}\log M_k\left(\tilde{E}_k,\tilde{F}_k,\tilde{G}_k,0\right) \\
&= \frac{1}{2K}\left[\sum_{c=1}^{N}\log\left(1+\frac{\alpha}{\sigma^2}(p_c-q_c)\right) + 2\alpha E_c m_c + \alpha(2n-1)F_c q_c + \alpha G_c p_c \right. \\
&\quad \left. + \frac{\sigma_0^2+\alpha(p_{0c}-2m_c+q_c)}{\sigma^2+\alpha(p_c-q_c)+n\sigma_0^2+n\alpha(p_{0c}-2m_c+q_c)}\right] \\
&\quad + \frac{1}{2K}\sum_{k=1}^{K}\log\left(1+\tilde{E}_k\right) - \frac{\tilde{E}_k^2+\tilde{F}_k}{1+\tilde{E}_k-n\tilde{E}_k^2-n\tilde{F}_k} \\
&\xrightarrow{n\to 0} \frac{1}{2K}\left[\sum_{c=1}^{N}\log\left(1+\frac{\alpha}{\sigma^2}(p_c-q_c)\right) + \frac{F_c}{E_c} + 2\alpha E_c m_c - \alpha F_c q_c + \alpha G_c p_c\right] \\
&\quad + \frac{1}{2K}\sum_{k=1}^{K}\log\left(1+\tilde{E}_k\right) - \frac{\tilde{E}_k^2+\tilde{F}_k}{1+\tilde{E}_k} \\
&= \frac{\mathrm{F}(\boldsymbol{x})}{K}.
\end{aligned}
$$

This is the final result for the mismatched detector. The six macroscopic parameters $E_c, F_c, G_c, m_c, q_c, p_c$ are implicitly given by the simultaneous solution of the system of equations (15) to (16) and (20) to (22) with the definitions (10) to (11) for all chip times c. This system of equations can only be solved numerically.

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

$$\frac{\mathrm{F}(\boldsymbol{x})}{K} = \frac{1}{2K}\left[\sum_{c=1}^{N} \log\left(1+\frac{\alpha}{\sigma_0^2}(p_{0c}-m_c)\right)+1+\alpha E_c m_c\right] + \frac{1}{2K}\sum_{k=1}^{K}\log\left(1+\tilde{E}_k\right)-\tilde{E}_k$$
$$= \frac{1}{2K}\left[\sum_{c=1}^{N} \sigma_0^2 E_c - \log\left(\sigma_0^2 E_c\right)\right] + \frac{1}{2K}\sum_{k=1}^{K}\log\left(1+\tilde{E}_k\right)$$

with

$$E_c = \frac{1}{\sigma_0^2 + \frac{\alpha}{K}\sum\limits_{k=1}^{K}\frac{w_{ck}^2}{1+\tilde{E}_k}}. \tag{24}$$

This result is more compact and it requires only to solve (24) numerically which is conveniently done by fixed-point iteration.

Comparing (24) to Girko's law, $\tilde{E}_k$ is recognized as the signal-to-interference and noise ratio of user k.

Using the similarity of free energy and the entropy of the channel output allows for the simple relationship

$$\frac{\mathrm{I}(\boldsymbol{x},\boldsymbol{y})}{K} = \frac{\mathrm{F}(\boldsymbol{x})}{K} - \frac{1}{2\alpha} \tag{25}$$

between the (normalized) free energy and the (normalized) mutual information between channel input signal $\boldsymbol{x}$ and channel output signal y given the channel matrix $\boldsymbol{H}$. Assuming that the channel is perfectly known to the receiver, but totally unknown to the transmitter, (25) gives the channel capacity per user.

Binary Prior Distribution

Consider a non-uniform binary prior

$$\mathrm{p}_a(x_{ak}) = \frac{1+t_k}{2}\,\delta(x_{ak}-1) + \frac{1-t_k}{2}\,\delta(x_{ak}+1). \tag{26}$$

Plugging the prior distribution into (12), 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 $t_k = \tanh(\lambda_k/2)$.

In the large system limit, the integral in (7) is dominated by that value of the integration variable which maximizes the argument of the exponential function. Thus, 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

$$m_c = \frac{1}{K} \sum_{k=1}^{K} w_{ck}^2 \int \tfrac{1+t_k}{2} \tanh\left(z\sqrt{\tilde{F}_k} + \tilde{E}_k + \tfrac{\lambda_k}{2}\right) + \tfrac{1-t_k}{2} \tanh\left(z\sqrt{\tilde{F}_k} + \tilde{E}_k - \tfrac{\lambda_k}{2}\right) \mathrm{D}z \tag{27}$$

$$q_c = \frac{1}{K} \sum_{k=1}^{K} w_{ck}^2 \int \tfrac{1+t_k}{2} \tanh^2\left(z\sqrt{\tilde{F}_k} + \tilde{E}_k + \tfrac{\lambda_k}{2}\right) + \tfrac{1-t_k}{2} \tanh^2\left(z\sqrt{\tilde{F}_k} + \tilde{E}_k - \tfrac{\lambda_k}{2}\right) \mathrm{D}z \tag{28}$$

$$p_c = p_{0c} = \frac{1}{K} \sum_{k=1}^{K} w_{ck}^2 \tag{29}$$

in the limit $n \to 0$.

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

$$
\begin{aligned}
-\frac{1}{K}\frac{\partial}{\partial n}\log \Xi_n &= \frac{1}{K}\frac{\partial}{\partial n}\sum_{c=1}^{N}\left[-\mathcal{G}(m_c,q_c,p_c,p_{0c}) + \alpha n E_c m_c + \frac{\alpha n(n-1)}{2}F_c q_c + \frac{\alpha n}{2}G_c p_c\right] \\
&\quad -\sum_{k=1}^{K}\log M_k\left(\tilde{E}_k,\tilde{F}_k,\tilde{G}_k,0\right) \\
&\stackrel{n\to 0}{\longrightarrow} \frac{1}{2K}\sum_{c=1}^{N}\left[\log\left(1+\frac{\alpha}{\sigma^2}(p_c-q_c)\right)+\frac{F_c}{E_c}+2\alpha E_c m_c - \alpha F_c q_c + \alpha G_c p_c\right] \\
&\quad -\frac{1}{K}\sum_{k=1}^{K}\int \frac{1+t_k}{2}\log\cosh\left(z\sqrt{\tilde{F}_k}+\tilde{E}_k+\frac{\lambda_k}{2}\right) \\
&\quad +\frac{1-t_k}{2}\log\cosh\left(z\sqrt{\tilde{F}_k}+\tilde{E}_k-\frac{\lambda_k}{2}\right)\mathrm{D}z + \frac{1}{2}\log\left(1-t_k^2\right)-\frac{\tilde{F}_k+\tilde{G}_k}{2} \\
&= \frac{\mathrm{F}(\boldsymbol{x})}{K}.
\end{aligned}
$$

This is the final result for the free energy of the mismatched detector. The six macroscopic parameters $E_c, F_c, G_c, m_c, q_c, p_c$ are implicitly given by the simultaneous solution of the system of equations (15) to (16) and (27) to (29) with the definitions (10) to (11) for all chip times c. 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.

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

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

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.

For any input alphabet to the channel mutual information is given by (25) with the free energy corresponding to that input alphabet.

MC-CDMA in Multipath Fading

Equivalent baseband vector channel in frequency domain:

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

$\boldsymbol{y}$	$\boldsymbol{W}$	$\odot$	$\boldsymbol{S}$	$\boldsymbol{x}$	$\boldsymbol{n}$
$N \times 1$	$N \times K$		$N \times K$	$K \times 1$	$N \times 1$
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)

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

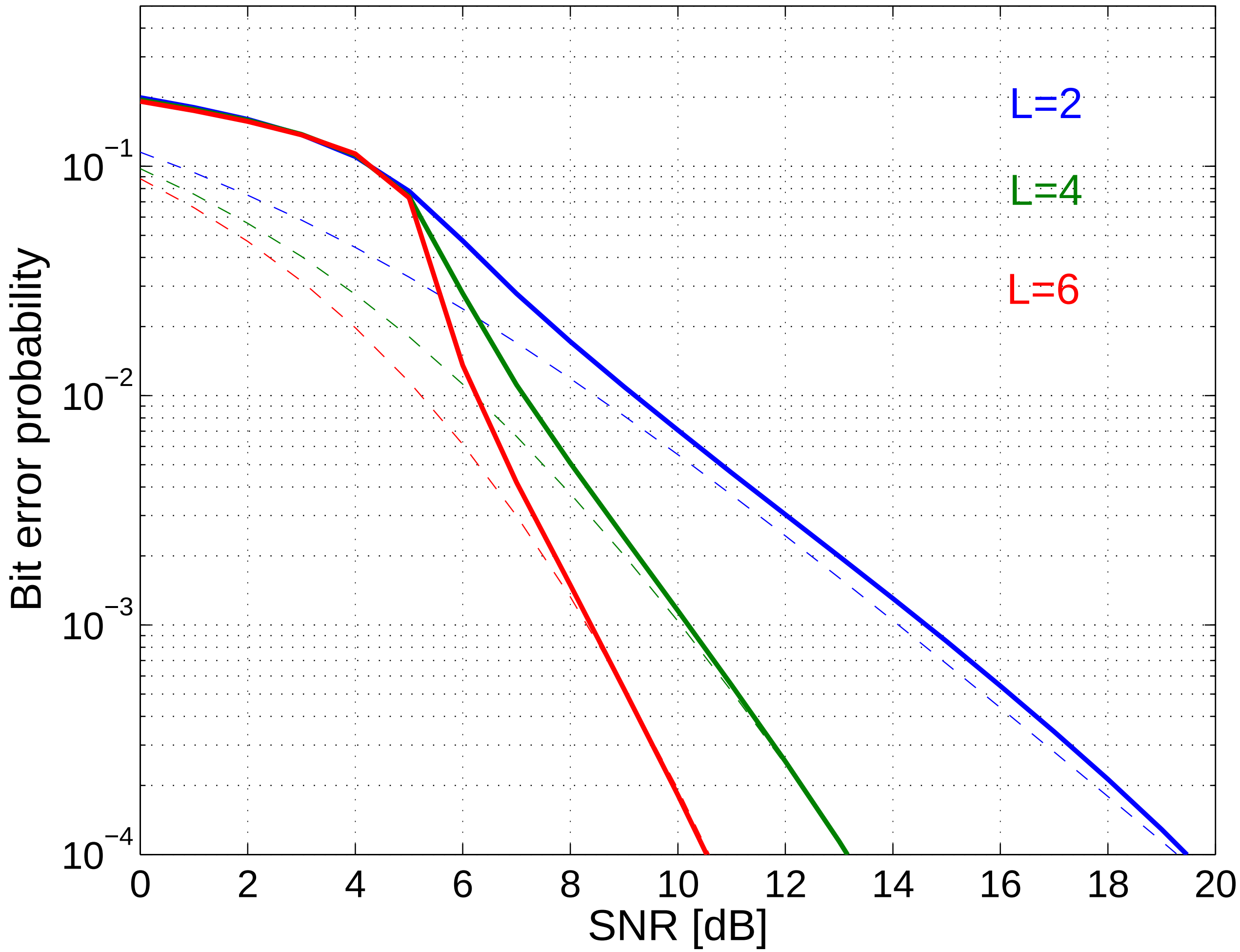
All users experience the same fading channel except for a scalar factor u_k.

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

Full diversity is achieved.

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

Uplink

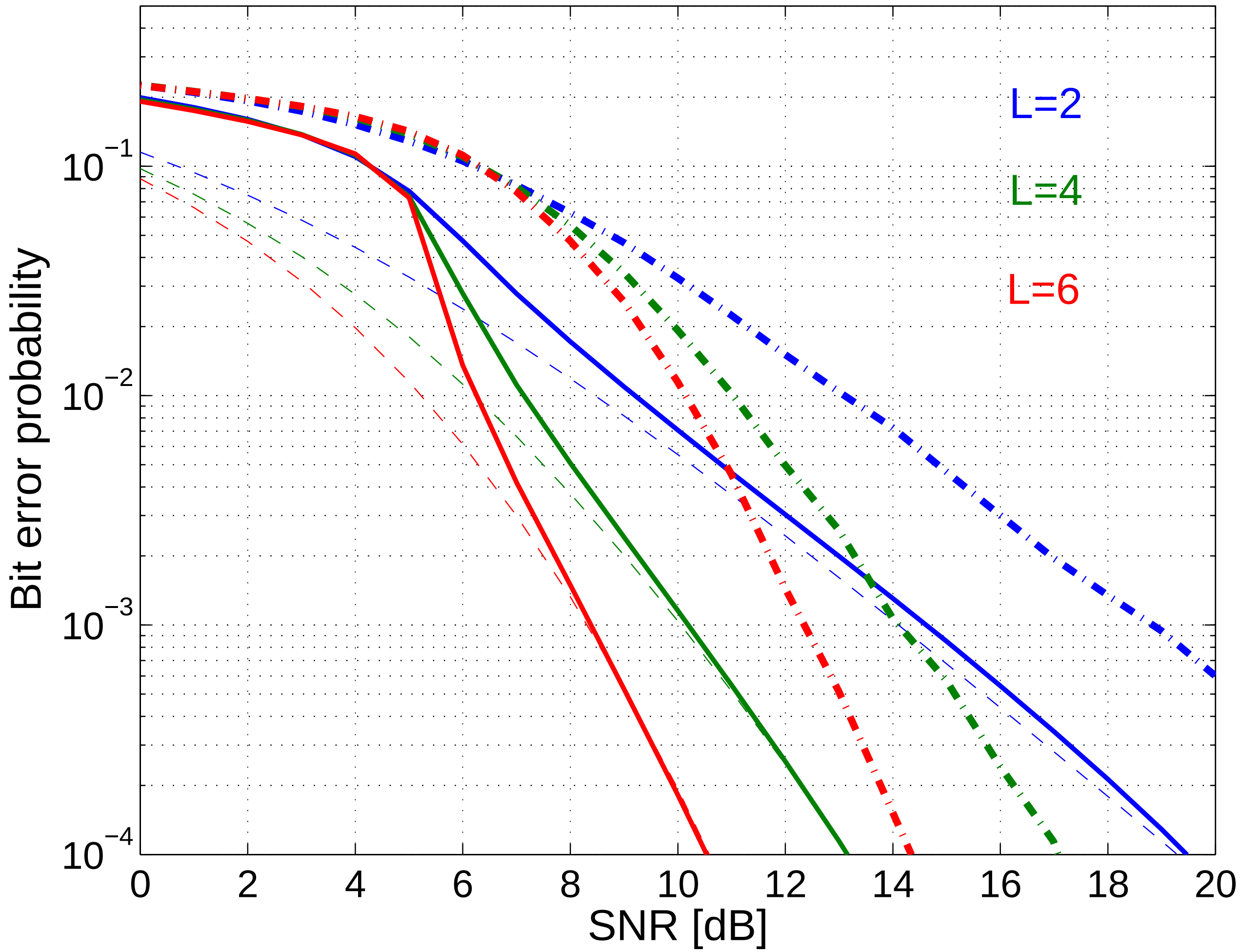


Uniform priors

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!

Minimization of a Quadratic Form

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

Minimization of a Quadratic Form

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:

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

Minimization of a Quadratic Form

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

Minimization of a Quadratic Form

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:

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

Minimization of a Quadratic Form

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

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

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

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

Replica Continuity

We want

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

Replica Continuity

We want

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

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

Replica Continuity

We want

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

Replica Continuity

We want

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

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

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

$$= \lim_{K\to\infty}\lim_{n\to 0}\frac{1}{nK}\log\operatorname*{E}_{\boldsymbol{Q}}\exp\left[-K\sum_{a=1}^{n}\int_0^{\lambda_a(\beta\boldsymbol{Q})}R_{\boldsymbol{J}}(-w)\,\mathrm{d}w\right]$$

with

$$Q_{ab} := \frac{1}{K}\boldsymbol{x}_a^{\mathrm{H}}\boldsymbol{x}_b = \frac{1}{K}\sum_{k=1}^{K}x_{ak}^{*}x_{bk}.$$

Let

$$\Xi_n = \lim_{K\to\infty} \frac{1}{K} \log \sum_{\{\boldsymbol{x}_a \in \mathcal{X}\}} \exp\left[-K \sum_{a=1}^{n} \int_0^{\lambda_a} R(-w)dw\right] \tag{30}$$

with $\lambda_1, \dots, \lambda_n$ denoting the eigenvalues of the $n \times n$ dimensional matrix $\beta\boldsymbol{Q}$,

In order to perform the summation in (30), the Kn-dimensional space spanned by the replicas is split into subshells

$$\mathcal{S}(\boldsymbol{Q}) \stackrel{\triangle}{=} \left\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n \,\middle|\, \boldsymbol{x}_a^{\mathrm{H}}\boldsymbol{x}_b = KQ_{ab}\right\} \tag{31}$$

where the inner product of two replicated vectors $\boldsymbol{x}_a$ and $\boldsymbol{x}_b$ is constant in each subshell. Noting that $\boldsymbol{x}_a^{\mathrm{H}}\boldsymbol{x}_b$ is Hermitian, we can express Ξ_n as

$$\Xi_n = \lim_{K\to\infty} \frac{1}{K} \log \int \mathrm{e}^{K\mathcal{I}(\boldsymbol{Q})} \mathrm{e}^{-K\mathcal{G}(\boldsymbol{Q})} \mathcal{D}\boldsymbol{Q}, \tag{32}$$

where

$$\mathcal{D}\boldsymbol{Q} = \prod_{a=1}^{n} \mathrm{d}Q_{aa} \prod_{b=a+1}^{n} \mathrm{d}\Re Q_{ab}\, \mathrm{d}\Im Q_{ab} \tag{33}$$

is the appropriate integration measure (note that $\boldsymbol{Q}^{\mathrm{H}} = \boldsymbol{Q}$),

$$\mathcal{G}(\boldsymbol{Q}) = \sum_{a=1}^{n} \int\limits_{0}^{\lambda_a(\beta \boldsymbol{Q})} R(-w)\,\mathrm{d}w \tag{34}$$

and

$$\begin{aligned} \mathrm{e}^{K\mathcal{I}(\boldsymbol{Q})} \;=\; & \sum_{\{\boldsymbol{x}_a \in \mathcal{B}_{\boldsymbol{s}}\}} \prod_{a=1}^{n} \delta\left(\boldsymbol{x}_a^{\mathrm{H}} \boldsymbol{x}_a - K Q_{aa}\right) \times \\ & \prod_{b=a+1}^{n} \delta\left(\Re\left[\boldsymbol{x}_a^{\mathrm{H}} \boldsymbol{x}_b - K Q_{ab}\right]\right) \delta\left(\Im\left[\boldsymbol{x}_a^{\mathrm{H}} \boldsymbol{x}_b - K Q_{ab}\right]\right) \end{aligned} \tag{35}$$

denotes the probability weight of the subshell. There are two reasons of following this procedure and introducing the new variables Q_{ab}. First, this allows us to explicitly sum over $\{\boldsymbol{x}_a\}$ as will be seen below. Secondly, we expect that for large K a single subshell will dominate Ξ_n, which will also be observed below. In the following the two exponential terms in (32) are evaluated separately.

We start with the evaluation of the measure $\mathrm{e}^{K\mathcal{I}(\boldsymbol{Q})}$.

For future convenience, we introduce the complex variables

$$\begin{array}{ll} \tilde{Q}_{ab}^{(\mathrm{I})} & 1 \le a \le b \le n \\ \tilde{Q}_{ab}^{(\mathrm{Q})} & 1 \le a < b \le n. \end{array}$$

We also define the matrix $\tilde{\boldsymbol{Q}}$ with elements

$$\tilde{Q}_{aa} = \tilde{Q}_{aa}^{(\mathrm{I})} \quad (36)$$

$$\tilde{Q}_{ab} = \frac{\tilde{Q}_{ab}^{(\mathrm{I})} - \mathrm{j}\tilde{Q}_{ab}^{(\mathrm{Q})}}{2} \quad (37)$$

$$\tilde{Q}_{ba} = \frac{\tilde{Q}_{ab}^{(\mathrm{I})} + \mathrm{j}\tilde{Q}_{ab}^{(\mathrm{Q})}}{2} \quad (38)$$

where $a < b$.

We may now write the Dirac measure of the elements of the Hermitian matrix

$$P_{ab} = \boldsymbol{x}_a^{\mathrm{H}} \boldsymbol{x}_b - K Q_{ab}$$

in terms of its inverse Laplace transform

$$\delta\left(P_{aa}\right) = \int\limits_{\mathcal{J}} \exp\left[\tilde{Q}_{aa} P_{aa}\right] \frac{\mathrm{d}\tilde{Q}_{aa}^{(\mathrm{I})}}{2\pi \mathrm{j}} \tag{39}$$

$$\delta\left(\Re P_{ab}\right) \delta\left(\Im P_{ab}\right) = \int\limits_{\mathcal{J}^2} \mathrm{e}^{\tilde{Q}_{ab}^{(\mathrm{I})} \Re P_{ab} - \tilde{Q}_{ab}^{(\mathrm{Q})} \Im P_{ab}} \frac{\mathrm{d}\tilde{Q}_{ab}^{(\mathrm{I})} \mathrm{d}\tilde{Q}_{ab}^{(\mathrm{Q})}}{(2\pi \mathrm{j})^2} \tag{40}$$

$$= \int\limits_{\mathcal{J}^2} \mathrm{e}^{\tilde{Q}_{ab} P_{ba} + \tilde{Q}_{ba} P_{ab}} \frac{\mathrm{d}\tilde{Q}_{ab}^{(\mathrm{I})} \mathrm{d}\tilde{Q}_{ab}^{(\mathrm{Q})}}{(2\pi \mathrm{j})^2}. \tag{41}$$

with $\mathcal{J} = (t - \mathrm{j}\infty; t + \mathrm{j}\infty)$ for some $t \in \mathbb{R}$ and $P_{ab} = P_{ba}^*$.

We may now express (35) as

$$\mathrm{e}^{K\mathcal{I}(\boldsymbol{Q})} = \sum_{\{\boldsymbol{x}_a \in \mathcal{B}_{\boldsymbol{s}}\}} \int_{\mathcal{J}^{n^2}} \mathrm{e}^{\sum_{a,b} \tilde{Q}_{ab}\left(\boldsymbol{x}_a^{\mathrm{H}}\boldsymbol{x}_b - KQ_{ab}\right)} \tilde{\mathcal{D}}\tilde{\boldsymbol{Q}} \tag{42}$$

$$= \int_{\mathcal{J}^{n^2}} \mathrm{e}^{-K\mathrm{tr}\left[\tilde{\boldsymbol{Q}}\boldsymbol{Q}\right] + \sum_{k=1}^{K} \log M_k\left(\tilde{\boldsymbol{Q}}\right)} \tilde{\mathcal{D}}\tilde{\boldsymbol{Q}} \tag{43}$$

where the integration measure is given by

$$\tilde{\mathcal{D}}\tilde{\boldsymbol{Q}} = \prod_{a=1}^{n} \left(\frac{\mathrm{d}\tilde{Q}_{aa}^{(I)}}{2\pi \mathrm{j}} \prod_{b=a+1}^{n} \frac{\mathrm{d}\tilde{Q}_{ab}^{(I)} \mathrm{d}\tilde{Q}_{ab}^{(Q)}}{(2\pi \mathrm{j})^2} \right) \tag{44}$$

and

$$M_k\left(\tilde{\boldsymbol{Q}}\right) = \sum_{\{x_a \in \mathcal{B}_k\}} \mathrm{e}^{\sum_{a,b} x_a^* x_b \tilde{Q}_{ab}}. \tag{45}$$

In the limit of $K \to \infty$ one of the exponential terms in (32) will dominate over all others. Thus, only that extremal value of the correlation Q_{ab} is relevant for calculation of the integral.

To make further progress, we need to identify the saddle-point which dominates the integrals. We invoke an important assumption on the structure of the matrices (Q_{ab}) and $(\tilde{Q}_{ab})$ at the saddle-point:

Assumption 1 (replica symmetry) *When applying the replica method to solve the saddle-point equations, we will assume that the extremal point is invariant to permutations of the replica indexes.*

The assumption of replica symmetry translates to searching over a subset of possible saddle-points with specific symmetry properties of the matrix $\boldsymbol{Q} = (Q_{ab})$. Indeed, we require that

$$\begin{aligned} Q_{ab} &= q \qquad \forall a \neq b \\ Q_{aa} &= q + \chi/\beta \qquad \forall a \end{aligned}$$

for some q and χ with $\chi \geq 0$ since $\boldsymbol{Q}$ has to be positive semidefinite. Thus we distinguish the correlation between different replicas and autocorrelation of an individual replica.

Replica Symmetry

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

with some macroscopic parameters q and χ.

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

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

Note that despite the fact that $\boldsymbol{Q}$ is complex-valued in general, its values at the saddle-point are in fact real-valued.

For the evaluation of $\mathcal{G}(\boldsymbol{Q})$ in (32), 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. Thus we get

$$\mathcal{G}(q, \chi) = (n-1) \int_0^{\chi} R(-w)\,\mathrm{d}w + \int_0^{\chi + \beta n q} R(-w)\,\mathrm{d}w. \qquad (46)$$

Since the integral in (32) is dominated by the maximum argument of the exponential function, the derivatives of

$$\mathcal{G}(q,\chi) + \mathrm{tr}(\tilde{\boldsymbol{Q}}\boldsymbol{Q}) \tag{47}$$

with respect to q and χ must vanish as $K \to \infty$.[3] The assumption of replica symmetry leads to

$$\mathrm{tr}(\tilde{\boldsymbol{Q}}\boldsymbol{Q}) = n(n-1)\beta^2 f^2 q + n\left(\beta f^2 - e\right)(\beta q + \chi)\,. \tag{48}$$

Taking derivatives after plugging (46) and (48) into (47) yields

$$nR(-\chi - \beta nq) + n(n-1)\beta f^2 + n\left(\beta f^2 - e\right) = 0 \tag{49}$$

$$(n-1)R(-\chi) + R(-\chi - \beta nq) + n\left(\beta f^2 - e\right) = 0 \tag{50}$$

and solving for e and f gives

$$e = R(-\chi) \tag{51}$$

$$f = \sqrt{\frac{R(-\chi) - R(-\chi - \beta nq)}{\beta n}} \quad \overset{n\to 0}{\longrightarrow} \quad \sqrt{qR'(-\chi)}. \tag{52}$$

[3] It turns out that when $\lim_{n\to 0} \partial_n \Xi_n$ is expressed in terms of e, f, q, χ, the relevant extremum is in fact a maximum and not a minimum. This is due to the fact that when n drops below unity, the minima of a function become maxima and vice-versa. For a detailed analysis of this technicality, see [11].

In addition, the replica symmetry assumption simplifies (45)

$$M_k(e,f) = \sum_{\{x_a\in\mathcal{B}_k\}} \mathrm{e}^{\beta\sum\limits_{a=1}^{n}\left[(\beta f^2-e)|x_a|^2+2\sum\limits_{b=a+1}^{n}\beta f^2\Re\{x_a^*x_b\}\right]} \tag{53}$$

$$= \sum_{\{x_a\in\mathcal{B}_k\}} \mathrm{e}^{\beta^2f^2\left|\sum\limits_{a=1}^{n}x_a\right|^2-\sum\limits_{a=1}^{n}\beta e|x_a|^2} \tag{54}$$

Note that the sets $\mathcal{B}_k$ enter the transmitted energy only via (54). Now, we apply the complex Hubbard-Stratonovich transform

$$\mathrm{e}^{|x|^2} = \int\limits_{\mathbb{C}} \mathrm{e}^{2\Re\{xz^*\}}\underbrace{\mathrm{e}^{-|z|^2}\frac{\mathrm{d}z}{\pi}}_{\triangleq\mathrm{D}z} \tag{55}$$

to (54).

With Hubbard-Stratonovich, we find

$$M_k(e,f) = \sum_{\{x_a \in \mathcal{B}_k\}} \int \mathrm{e}^{\beta \sum\limits_{a=1}^{n} 2f\Re\{x_a z^*\} - e|x_a|^2} \mathrm{D}z \tag{56}$$

$$= \int \left(\sum_{x \in \mathcal{B}_k} \mathrm{e}^{2\beta f \Re\{x z^*\} - \beta e |x|^2} \right)^n \mathrm{D}z. \tag{57}$$

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

$$\log M(e,f) = \frac{1}{K} \sum_{k=1}^{K} \log M_k(e,f) \tag{58}$$

$$\to \int \log \int \left(\sum_{x \in \mathcal{B}} \mathrm{e}^{2\beta f \Re\{z^* x\} - \beta e |x|^2} \right)^n \mathrm{D}z \mathrm{dP}(\mathcal{B}). \tag{59}$$

In the large-system limit, the integral in (43) is dominated by that value of the integration variable which maximizes the exponent. Thus, partial derivatives of

$$\log M(e,f) - \operatorname{tr}(\tilde{\boldsymbol{Q}}\boldsymbol{Q}) \tag{60}$$

with respect to f and e must vanish as $K \to \infty$. An explicit calculation of the two derivatives gives the following expressions for the macroscopic parameters q and χ

$$\chi = \frac{1}{\sqrt{qR'(-\chi)}} \iint \frac{\sum_{x\in\mathcal{B}} \Re\{z^* x\} \mathrm{e}^{\beta 2\sqrt{qR'(-\chi)}\Re\{z^*x\} - \beta R(-\chi)|x|^2}}{\sum_{x\in\mathcal{B}} \mathrm{e}^{\beta 2\sqrt{qR'(-\chi)}\Re\{z^*x\} - \beta R(-\chi)|x|^2}} \mathrm{D}z\, \mathrm{dP}(\mathcal{B}) \tag{61}$$

$$q = \iint \frac{\sum_{x\in\mathcal{B}} |x|^2 \mathrm{e}^{\beta 2\sqrt{qR'(-\chi)}\Re\{z^*x\} - \beta R(-\chi)|x|^2}}{\sum_{x\in\mathcal{B}} \mathrm{e}^{\beta 2\sqrt{qR'(-\chi)}\Re\{z^*x\} - \beta R(-\chi)|x|^2}} \mathrm{D}z\, \mathrm{dP}(\mathcal{B}) - \frac{\chi}{\beta}. \tag{62}$$

Finally, the fixed-point equations (61) and (62) simplify via the saddle point integration rule to

$$\chi = \frac{1}{\sqrt{qR'(-\chi)}} \iint \Re \operatorname*{argmin}_{x\in\mathcal{B}} \left| z - \frac{R(-\chi)\, x}{\sqrt{qR'(-\chi)}} \right| z^* \mathrm{D}z \, \mathrm{dP}(\mathcal{B}) \tag{63}$$

$$q = \iint \left| \operatorname*{argmin}_{x\in\mathcal{B}} \left| z - \frac{R(-\chi)\, x}{\sqrt{qR'(-\chi)}} \right| \right|^2 \mathrm{D}z \, \mathrm{dP}(\mathcal{B}) \tag{64}$$

in the limit $\beta \to \infty$. Note that the minimization with respect to the symbol x splits the integration space of z into the Voronoi regions defined by the (appropriately scaled) signal constellation $\mathcal{B}$.

Returning to the initial goal of the minization of the quadratic form, and collecting previous results, we find with replica continuity that

$$E = \lim_{\beta\to\infty} \frac{1}{\beta} \lim_{n\to 0} \frac{\partial}{\partial n} \Bigg[(n-1) \int\limits_0^{\chi} R(-w)\,\mathrm{d}w + \int\limits_0^{\chi+\beta n q} R(-w)\,\mathrm{d}w - \log M(e,f) + n(n-1) f^2 \beta^2 q + n(f^2\beta - e)(\chi + \beta q) \Bigg] \quad (65)$$

$$= \lim_{\beta\to\infty} \frac{1}{\beta} \int\limits_0^{\chi} R(-w)\,\mathrm{d}w - \frac{\chi}{\beta} R(-\chi) + q\chi R'(-\chi) - \frac{1}{\beta} \iint \log \sum_{x\in\mathcal{B}} \mathrm{e}^{\beta 2 f \Re\{z^* x\} - \beta e |x|^2} \mathrm{D}z \, \mathrm{dP}(\mathcal{B}). \quad (66)$$

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

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

Note that for any bound on the amplitude of the signal set $\mathcal{B}$, the parameter q is finite. Even without bound, q will remain finite for a well-defined minimization problem. The parameter χ behaves in a more complicated manner. It can be both zero, finite, and infinite as $\beta \to \infty$ depending on the particular R-transform and the signal sets $\mathcal{B}_s$. For $\chi \notin (0, \infty)$, the saddle-point limits have to be reconsidered.

1st Order Replica Symmetry Breaking (1RSB)

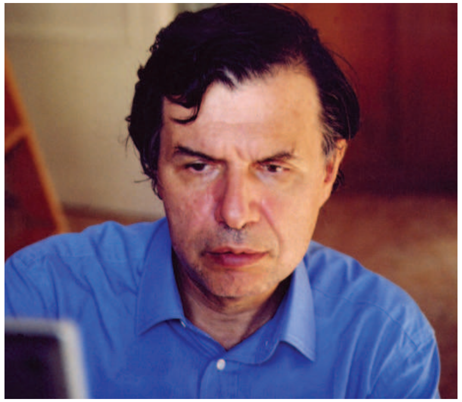
$$
\boldsymbol{Q} := \begin{bmatrix}
\overbrace{q+p+\frac{\chi}{\beta} \quad q+p}^{\frac{\mu}{\beta}\text{ columns}} & q & q & \cdots & q & q \\
q+p \quad q+p+\frac{\chi}{\beta} & q & q & \cdots & q & q \\
q \quad\quad q & q+p+\frac{\chi}{\beta} & q+p & \ddots & q & q \\
q \quad\quad q & q+p & q+p+\frac{\chi}{\beta} & & \vdots & \vdots \\
\vdots \quad\quad \vdots & \ddots & & \ddots & q & q \\
q \quad\quad q & q & \cdots & q & q+p+\frac{\chi}{\beta} & q+p \\
q \quad\quad q & q & \cdots & q & q+p & q+p+\frac{\chi}{\beta}
\end{bmatrix}
$$

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

Higher order 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.