The Decoupling Principle
for High-Dimensional Gaussian Vector Channels

Ralf R. Müller

Friedrich-Alexander Universität Erlangen-Nürnberg

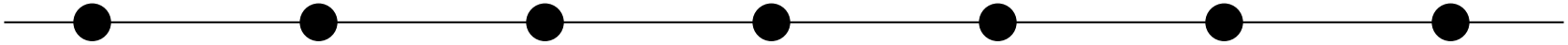
ralf.r.mueller@fau.de

March 2, 2026

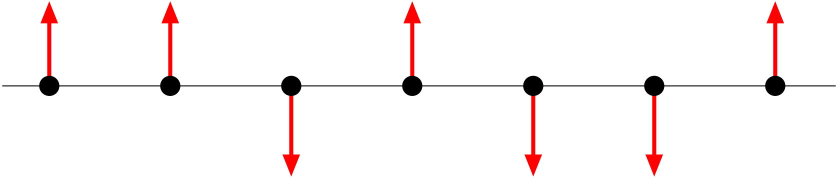
Table of Contents

Table of Contents

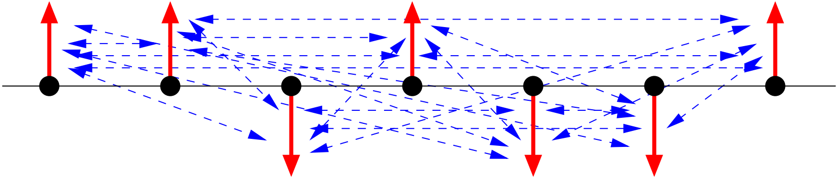
Magnetic Interactions



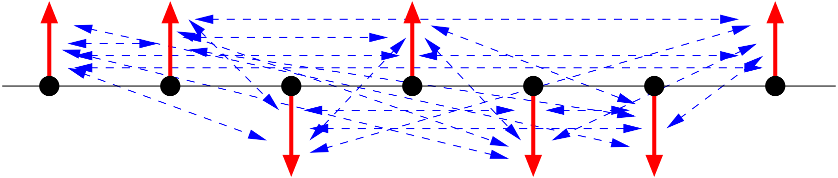
Magnetic Interactions



Magnetic Interactions



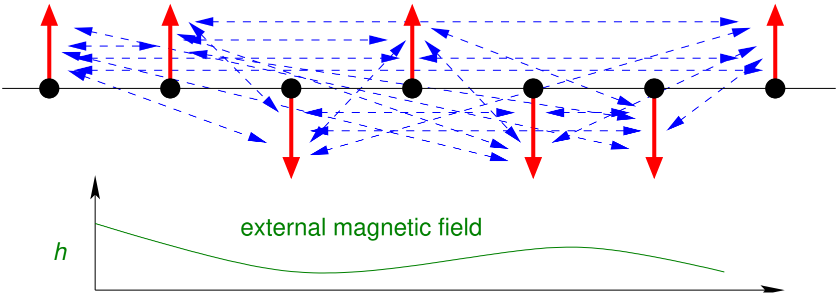
Magnetic Interactions



Energy function (Hamiltonian):

$$\sum_k \sum_{i \neq k} r_{ki} x_k x_i$$

Magnetic Interactions



Energy function (Hamiltonian):

$$\sum_{k} \sum_{i \neq k} r_{ki} x_k x_i + \sum_{k} h_k x_k$$

Optimal Detection of Vector Channel Observation

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n} \qquad \in \mathbb{R}^N$$

Optimal Detection of Vector Channel Observation

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n} \qquad \in \mathbb{R}^N$$

Best estimate for transmitted data:

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

Optimal Detection of Vector Channel Observation

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n} \qquad \in \mathbb{R}^N$$

Best estimate for transmitted data:

$$\begin{aligned}\hat{\boldsymbol{x}} &= \underset{\boldsymbol{x}\in\{\pm 1\}^K}{\operatorname{argmin}} ||\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}|| \\ &= \underset{\boldsymbol{x}\in\{\pm 1\}^K}{\operatorname{argmin}} \boldsymbol{x}^\dagger \boldsymbol{R}\boldsymbol{x} + \boldsymbol{h}^\dagger \boldsymbol{x} + \boldsymbol{y}^\dagger \boldsymbol{y} \qquad \text{with} \qquad \begin{aligned}\boldsymbol{R} &= \boldsymbol{H}^\dagger \boldsymbol{H} \\ \boldsymbol{h} &= -2\boldsymbol{H}^\dagger \boldsymbol{y}\end{aligned}\end{aligned}$$

Optimal Detection of Vector Channel Observation

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n} \qquad \in \mathbb{R}^N$$

Best estimate for transmitted data:

$$\begin{aligned}
\hat{\boldsymbol{x}} &= \underset{\boldsymbol{x}\in\{\pm 1\}^K}{\operatorname{argmin}} \, ||\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}|| \\
&= \underset{\boldsymbol{x}\in\{\pm 1\}^K}{\operatorname{argmin}} \, \boldsymbol{x}^\dagger \boldsymbol{R} \boldsymbol{x} + \boldsymbol{h}^\dagger \boldsymbol{x} \qquad \text{with} \qquad \begin{array}{rcl} \boldsymbol{R} &=& \boldsymbol{H}^\dagger \boldsymbol{H} \\ \boldsymbol{h} &=& -2\boldsymbol{H}^\dagger \boldsymbol{y} \end{array} \\
&= \underset{\boldsymbol{x}\in\{\pm 1\}^K}{\operatorname{argmin}} \sum_k \sum_{i\neq k} r_{ki} x_k x_i + \sum_k h_k x_k + \sum_k r_{kk} x_k^2
\end{aligned}$$

Optimal Detection of Vector Channel Observation

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n} \qquad \in \mathbb{R}^N$$

Best estimate for transmitted data:

$$\begin{aligned} \hat{\boldsymbol{x}} &= \underset{\boldsymbol{x}\in\{\pm 1\}^K}{\operatorname{argmin}} \, ||\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}|| \\ &= \underset{\boldsymbol{x}\in\{\pm 1\}^K}{\operatorname{argmin}} \, \boldsymbol{x}^\dagger \boldsymbol{R}\boldsymbol{x} + \boldsymbol{h}^\dagger \boldsymbol{x} \qquad \text{with} \qquad \begin{aligned} \boldsymbol{R} &= \boldsymbol{H}^\dagger \boldsymbol{H} \\ \boldsymbol{h} &= -2\boldsymbol{H}^\dagger \boldsymbol{y} \end{aligned} \\ &= \underset{\boldsymbol{x}\in\{\pm 1\}^K}{\operatorname{argmin}} \sum_k \sum_{i\neq k} r_{ki} x_k x_i + \sum_k h_k x_k \end{aligned}$$

Minimization of the energy function in physics!

The Obvious Decoupling Principle

Consider a very simplistic estimator (commonly termed matched filter receiver)

$$\hat{\boldsymbol{x}} = \operatorname{sign}\left(\boldsymbol{H}^\dagger \boldsymbol{y}\right)$$

i.e.

$$\hat{x}_k = \operatorname{sign}\left(r_{kk}x_k + \sum_{i\neq k} r_{ki}x_i + \boldsymbol{h}_k^\dagger \boldsymbol{n}\right)$$

The Obvious Decoupling Principle

Consider a very simplistic estimator (commonly termed matched filter receiver)

$$\hat{\boldsymbol{x}} = \operatorname{sign}\left(\boldsymbol{H}^\dagger \boldsymbol{y}\right)$$

i.e.

$$\hat{x}_k = \operatorname{sign}\left(r_{kk}x_k + \sum_{i\neq k} r_{ki}x_i + \boldsymbol{h}_k^\dagger \boldsymbol{n}\right)$$

$$\mathrm{SINR}_k = \frac{r_{kk}^2}{\text{noise power} + \sum_{i\neq k} r_{ki}^2}$$

- The contributions to interference power are additive.
- The interference power is subject to the law of large numbers.

∃ equivalent additive noise channel with noise power increased by interference power.

The Linear MMSE Decoupling Principle

Consider a linear MMSE estimator

$$\hat{\boldsymbol{x}} = \operatorname{sign}\left(\left(\boldsymbol{H}^\dagger \boldsymbol{H} + \sigma_n^2 \mathbf{I}\right)^{-1} \boldsymbol{H}^\dagger \boldsymbol{y}\right)$$

For $H_{ij} \sim \mathcal{N}(0, 1/N)$, Tse & Hanly found in 1997

$$\mathrm{SINR}_k \approx \frac{r_{kk}^2}{\sigma_n^2 + \underbrace{\sum_{\substack{i=1 \\ i \neq k}} \frac{r_{ki}^2}{1 + \mathrm{SINR}_i}}_{\longrightarrow \sigma_I^2}}$$

- The formula becomes exact as $N, K \to \infty$ with fixed ratio $\alpha = K/N$.
- The contributions to interference power are still additive despite dependent interference terms.

$\exists$ equivalent additive noise channel with noise power increased by interference power σ_I^2.

The Linear MMSE Decoupling Principle

Consider a linear MMSE estimator

$$\hat{\boldsymbol{x}} = \operatorname{sign}\left(\left(\boldsymbol{H}^\dagger \boldsymbol{H} + \sigma_n^2 \mathbf{I}\right)^{-1} \boldsymbol{H}^\dagger \boldsymbol{y}\right)$$

For $H_{ij} \sim \mathcal{N}(0, 1/N)$, Tse & Hanly found in 1997

$$\mathrm{SINR}_k \approx \frac{1}{\sigma_n^2 + \underbrace{\sum_{\substack{i=1 \\ i \neq k}} \frac{1/N}{1 + \mathrm{SINR}_i}}_{\longrightarrow \sigma_I^2}}$$

- The formula becomes exact as $N, K \to \infty$ with fixed ratio $\alpha = K/N$.
- The contributions to interference power are still additive despite dependent interference terms.

$\exists$ equivalent additive noise channel with noise power increased by interference power σ_I^2.

The Linear MMSE Decoupling Principle

Consider a linear MMSE estimator

$$\hat{\boldsymbol{x}} = \operatorname{sign}\left(\left(\boldsymbol{H}^\dagger \boldsymbol{H} + \sigma_n^2 \mathbf{I}\right)^{-1} \boldsymbol{H}^\dagger \boldsymbol{y}\right)$$

For $H_{ij} \sim \mathcal{N}(0, 1/N)$, Tse & Hanly found in 1997

$$\mathrm{SINR}_k \approx \frac{1}{\sigma_n^2 + \underbrace{\sum_{\substack{i=1\\ i\neq k}} \frac{1/N}{1+\mathrm{SINR}_i}}_{\longrightarrow \sigma_I^2}}$$

- The formula becomes exact as $N, K \to \infty$ with fixed ratio $\alpha = K/N$.
- The contributions to interference power are still additive despite dependent interference terms.

$\exists$ equivalent additive noise channel with noise power increased by interference power σ_I^2.

WHY?

Freeness of Two Random Matrices

Theorem 1 (Voiculescu 1986)

Two random matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{K \times K}$ *are asymptotically free, if for all integers* $n, m > 0$*, and all* $\alpha_{ij}, \beta_{ij} \in \mathbb{R}$ *we have*

$$Tr\left(\prod_{j=1}^{n}\left[\sum_{i=0}^{m}\alpha_{ij}\left(\boldsymbol{A}^i - \mathbf{I}\,Tr\boldsymbol{A}^i\right)\right]\left[\sum_{i=0}^{m}\beta_{ij}\left(\boldsymbol{B}^i - \mathbf{I}\,Tr\boldsymbol{B}^i\right)\right]\right) = 0$$

where

$$Tr(\cdot) = \lim_{K\to\infty}\frac{1}{K}trace(\cdot).$$

Additive Free Convolution

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

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

Then,

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

Additive Free Convolution

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

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

Then,

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

The R-transform is defined as

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

where

$$\mathrm{G}_{\boldsymbol{X}}(w) = -\sum_{n=0}^{\infty} w^{-n-1} \mathrm{Tr} \boldsymbol{X}^n$$

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

Effective Interference (Tse 1999)

$$\mathrm{SINR}_k \approx \frac{1}{\sigma_n^2 + \frac{1}{K}\sum_{\substack{i=1\\ i\neq k}}^{K} \frac{\alpha}{1+\mathrm{SINR}_i}}$$

Effective interference from user i:

$$I_i = \frac{1}{K} \cdot \frac{\alpha}{1+\mathrm{SINR}_i}$$

Effective Interference (Tse 1999)

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

Effective interference from user i:

$$I_i = \frac{1}{K}\cdot\frac{\alpha}{1+\mathsf{SINR}_i}$$

Definition of R-transform:

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

R-transform for i.i.d. $\boldsymbol{H}$:

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

Effective Interference (Tse 1999)

$$\mathrm{SINR}_k \approx \frac{1}{\sigma_n^2 + \frac{1}{K}\sum_{\substack{i=1\\ i\neq k}}^{K} \frac{\alpha}{1+\mathrm{SINR}_i}}$$

Effective interference from user i:

$$I_i = \frac{1}{K}\cdot\frac{\alpha}{1+\mathrm{SINR}_i}$$

Definition of R-transform:

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

R-transform for i.i.d. $\boldsymbol{H}$:

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

The effective interferences are the R-transforms of all individual users.

The SINRs are the Stieltjes transforms at $w=-\sigma_n^2$

The Linear MMSE Decoupling Principle (cont'd)

In the large system limit, to the linear vector channel

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

with independent additive Gaussian noise of covariance $\sigma^2\mathbf{I}$, i.i.d. data $\boldsymbol{x} \in \{\pm 1\}^K$, i.i.d. random matrix $\boldsymbol{H}$, and linear MMSE detection $\boldsymbol{d} = f(\boldsymbol{y}|\boldsymbol{H})$, there exists a scalar channel

$$y = x + n$$

with independent additive Gaussian noise of variance σ^2/η, $\eta \le 1$, data $x \in \{\pm 1\}$, and linear MMSE detection $d = g(y)$ which has the same bit error rate.

The Linear MMSE Decoupling Principle (cont'd)

In the large system limit, to the linear vector channel

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

with independent additive Gaussian noise of covariance $\sigma^2\mathbf{I}$, i.i.d. data $\boldsymbol{x} \in \{\pm 1\}^K$, i.i.d. random matrix $\boldsymbol{H}$, and linear MMSE detection $\boldsymbol{d} = f(\boldsymbol{y}|\boldsymbol{H})$, there exists a scalar channel

$$y = x + n$$

with independent additive Gaussian noise of variance σ^2/η, $\eta \leq 1$, data $x \in \{\pm 1\}$, and linear MMSE detection $d = g(y)$ such that $\Pr(d_k, x_k) = \Pr(d, x), \forall k$.

The Linear MMSE Decoupling Principle (cont'd)

In the large system limit, to the linear vector channel

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

with independent additive Gaussian noise of covariance $\sigma^2\mathbf{I}$, i.i.d. data $\boldsymbol{x} \in \mathcal{X}^K$, i.i.d. random matrix $\boldsymbol{H}$, and linear MMSE detection $\boldsymbol{d} = f(\boldsymbol{y}|\boldsymbol{H})$, there exists a scalar channel

$$y = x + n$$

with independent additive Gaussian noise of variance σ^2/η, $\eta \leq 1$, data $x \in \mathcal{X}$, and linear MMSE detection $d = g(y)$ such that $\Pr(d_k, x_k) = \Pr(d, x), \forall k$.

The Linear MMSE Decoupling Principle (cont'd)

In the large system limit, to the linear vector channel

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

with independent additive Gaussian noise of covariance $\sigma^2 \mathbf{I}$, i.i.d. data $\boldsymbol{x} \in \mathcal{X}^K$, bi-unitarily invariant random matrix $\boldsymbol{H}$, and linear MMSE detection $\boldsymbol{d} = f(\boldsymbol{y}|\boldsymbol{H})$, there exists a scalar channel

$$y = x + n$$

with independent additive Gaussian noise of variance σ^2/η, $\eta \leq 1$, data $x \in \mathcal{X}$, and linear MMSE detection $d = g(y)$ such that $\Pr(d_k, x_k) = \Pr(d, x), \forall k$.

In general, the scale factor η depends on the asymptotic singular value distribution of $\boldsymbol{H}$, the noise variance σ^2, and the alphabet $\mathcal{X}$ of the data.

Table of Contents

The MAP Decoupling Principle (Tanaka, 2001)

In the large system limit, to the linear vector channel

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

with independent additive Gaussian noise of covariance $\sigma^2\mathbf{I}$, i.i.d. data $\boldsymbol{x} \in \{\pm 1\}^K$, i.i.d. random matrix $\boldsymbol{H}$, and component-wise MAP detection $\boldsymbol{d} = f(\boldsymbol{y}|\boldsymbol{H})$, there exists a scalar channel

$$y = x + n$$

with independent additive Gaussian noise of variance σ^2/η, $\eta \leq 1$, data $x \in \{\pm 1\}$, and MAP detection $d = g(y)$ which has the same bit error rate.

In general, the scale factor η depends on the asymptotic singular value distribution of $\boldsymbol{H}$, and the alphabet of the data.

The MAP Decoupling Principle (Tanaka, 2001)

In the large system limit, to the linear vector channel

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

with independent additive Gaussian noise of covariance $\sigma^2 \mathbf{I}$, i.i.d. data $\boldsymbol{x} \in \{\pm 1\}^K$, i.i.d. random matrix $\boldsymbol{H}$, and component-wise MAP detection $\boldsymbol{d} = f(\boldsymbol{y}|\boldsymbol{H})$, there exists a scalar channel

$$y = x + n$$

with independent additive Gaussian noise of variance σ^2/η, $\eta \leq 1$, data $x \in \{\pm 1\}$, and MAP detection $d = g(y)$ such that $\Pr(d_k, x_k) = \Pr(d, x), \forall k$.

In general, the scale factor η depends on the asymptotic singular value distribution of $\boldsymbol{H}$, and the alphabet of the data.

The RS Decoupling Principle

In the large system limit, to the linear vector channel

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

with independent additive Gaussian noise of covariance $\sigma^2\mathbf{I}$, i.i.d. data $\boldsymbol{x} \in \mathcal{X}^K$, bi-unitraily invariant random matrix $\boldsymbol{H}$, and appropriate estimation $\boldsymbol{d} = f(\boldsymbol{y}|\boldsymbol{H})$, there exists a scalar channel

$$y = x + n$$

with independent additive Gaussian noise of variance σ^2/η, $\eta \leq 1$, data $x \in \mathcal{X}$, and appropriate estimation $d = g(y)$ such that $\Pr(d_k, x_k) = \Pr(d, x), \forall k$, if **replica symmetry** holds.

In general, the scale factor η depends on the asymptotic singular value distribution of $\boldsymbol{H}$, the estimator $f(\cdot)$, and the alphabet of the data.

Decoupling Principle for Broken Replica Symmetry (2016)

In the large system limit, to the linear vector channel

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

with independent additive Gaussian noise of covariance $\sigma^2\mathbf{I}$, i.i.d. data $\boldsymbol{x} \in \mathcal{X}^K$, bi-unitarily invariant random matrix $\boldsymbol{H}$, and appropriate estimation $\boldsymbol{d} = f(\boldsymbol{y}|\boldsymbol{H})$, there exists a scalar channel

$$y = x + n$$

with dependent additive non-Gaussian noise, data $x \in \mathcal{X}$, and appropriate estimation $d = g(y)$ such that $\Pr(d_k, x_k) = \Pr(d, x), \forall k$, if **replica symmetry is broken**.

Energy vs. Entropy

The following two optimization problems are equivalent:

- Maximize the entropy $H(X)$ for fixed energy $E(X)$
- Minimize the energy $E(X)$ for fixed entropy $H(X)$

with

$$E(X) = \operatorname*{E}_{\boldsymbol{x}} ||\boldsymbol{x}||, \qquad H(X) = \operatorname*{E}_{\boldsymbol{x}} \log \Pr(\boldsymbol{x})$$

Define the free energy (average Lagrange function)

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

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

Equivalently, we have

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

Free Energy

With the Boltzmann distribution, the free energy is given by

$$F(X) = -\frac{1}{\beta} \log \left[\sum_i \mathrm{e}^{-\beta ||\boldsymbol{x}_i||} \right]$$

with $\beta = 1/T$ denoting inverse temperature. It depends only on the partition function.

Energy and entropy are found from the free energy simply as

$$E(X) = \frac{\partial}{\partial \beta} \beta F(X)$$

$$H(X) = \beta^2 \frac{\partial}{\partial \beta} F(X).$$

The Configurational Average

When analyzing a random system, we evaluate the free energy averaged over the random configuration of the system.

The random configuration is, e.g., the position of the electrons in a spin glass, the channel matrix in a multi-antenna channel, etc.

In information theory, that corresponds to Shannon's analysis of the average performance of all codes.

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.

According to the second law of thermodynamics and the law of energy conservation, nature maximizes entropy for a given energy.

In communications engineering, the energy function is the metric used by the decoder. The decoder does a job equivalent to nature, to minimize the metric for a given output entropy.

By defining an energy function, you implicitly define the forces ruling your thermodynamic system. These forces need not be in line with true physical forces. By contrast, defining an energy function, you can define your own toy universe.

Mismatched Energy Functions

The energy function is defined by the metric optimized in the decoder (communications) or estimator (signal processing). This metric is arbitrary. It need not be in line with the optimal metric or the true statistical distribution of the channel.

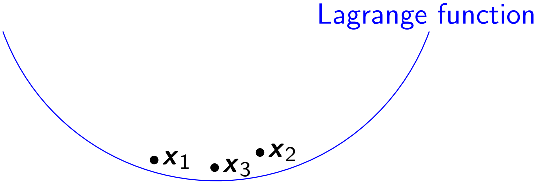
In such a case, the free energy is defined by the assumed (eventually suboptimal) metric of the decoder/estimator. However, the configurational average is with respect to the true statistical distribution.

This allows to analyze cases, where the estimator/decoder uses a wrong metric due to whatever reason, e.g. insufficient knowledge of the channel, complexity considerations, etc.

Table of Contents

Quenched Random Variables

The configurational average $\mathrm{E}_{\{\boldsymbol{x}_k\}}$ is substituted by an average over the quenched random variables $\boldsymbol{Q}$.

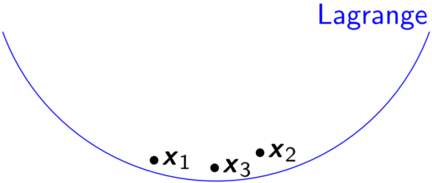


$$\boldsymbol{Q} = \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix}$$

with $q_{ki} = \langle \boldsymbol{x}_k ; \boldsymbol{x}_i \rangle$.

Quenched Random Variables

The configurational average $\mathrm{E}_{\{\boldsymbol{x}_k\}}$ is substituted by an average over the quenched random variables $\boldsymbol{Q}$.



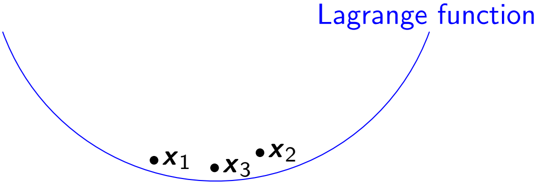
$$\boldsymbol{Q} = \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix}$$

with $q_{ki} = \langle \boldsymbol{x}_k ; \boldsymbol{x}_i \rangle$.

Consider $K \to \infty$ in the vicinity of the minimum of the objective function: If all off-diagonal elements of $\boldsymbol{Q}$ converge to the same value, we call the problem **replica symmetric**.

Quenched Random Variables

The configurational average $\mathrm{E}_{\{\boldsymbol{x}_k\}}$ is substituted by an average over the quenched random variables $\boldsymbol{Q}$.



$$\boldsymbol{Q} = \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix}$$

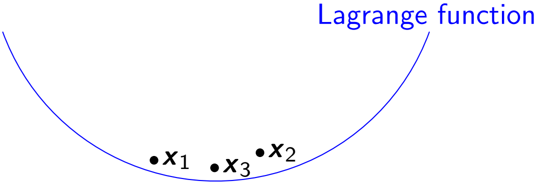
with $q_{ki} = \langle \boldsymbol{x}_k ; \boldsymbol{x}_i \rangle$.

Consider $K \to \infty$ in the vicinity of the minimum of the objective function: If all off-diagonal elements of $\boldsymbol{Q}$ converge to the same value, we call the problem **replica symmetric**.

- All convex problems are replica symmetric.

Quenched Random Variables

The configurational average $\mathrm{E}_{\{\boldsymbol{x}_k\}}$ is substituted by an average over the quenched random variables $\boldsymbol{Q}$.



$$\boldsymbol{Q} = \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix}$$

with $q_{ki} = \langle \boldsymbol{x}_k ; \boldsymbol{x}_i \rangle$.

Consider $K \to \infty$ in the vicinity of the minimum of the objective function: If all off-diagonal elements of $\boldsymbol{Q}$ converge to the same value, we call the problem **replica symmetric**.

- All convex problems are replica symmetric.
- Some NP-complete problems are replica symmetric, too.

The Meaning of Replica Symmetry Breaking

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

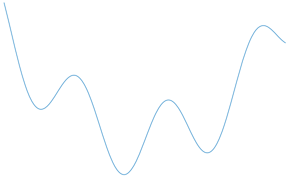
- If there are multiple minima, many of those are quite close to each other.
- The problem can often be well approximated by iterative algorithms like message passing.

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

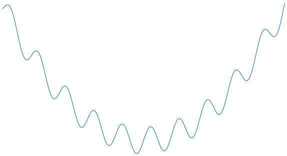
- There are extrema at very different positions.
- Message passing is often significantly suboptimum.

RS (breaking) ranks the difficulty of NP-complete problems, in practice.

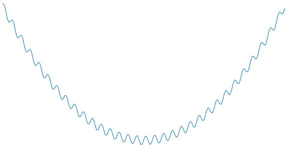
The Asymptotics of the Energy Function



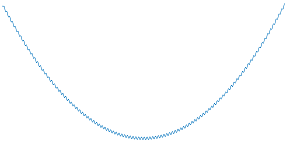
The Asymptotics of the Energy Function



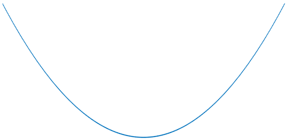
The Asymptotics of the Energy Function



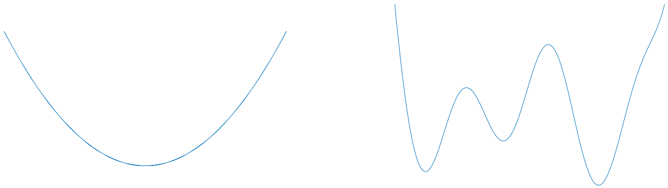
The Asymptotics of the Energy Function



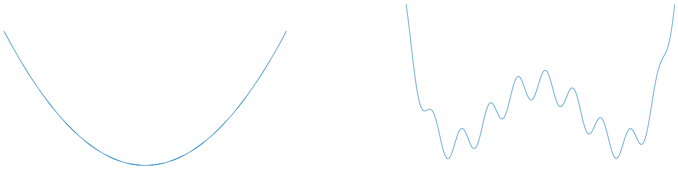
The Asymptotics of the Energy Function



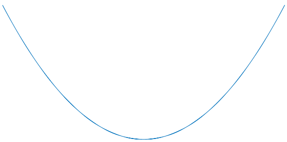
The Asymptotics of the Energy Function

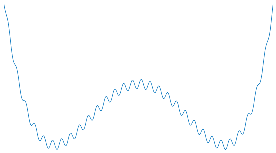


The Asymptotics of the Energy Function

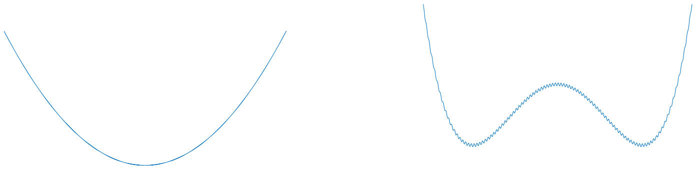


The Asymptotics of the Energy Function

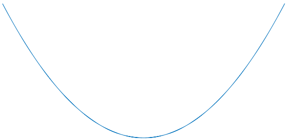


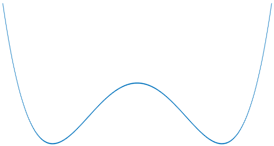


The Asymptotics of the Energy Function

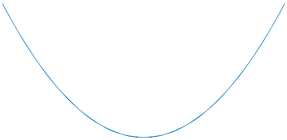


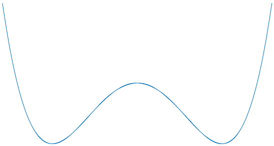
The Asymptotics of the Energy Function



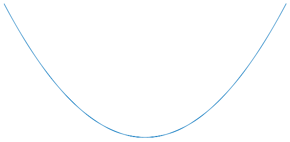


The Asymptotics of the Energy Function

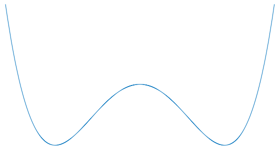




The Asymptotics of the Energy Function

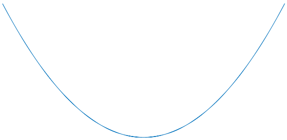


1 off-diagonal value (RS)

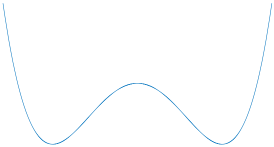


2 different off-diagonal values (1RSB)

The Asymptotics of the Energy Function



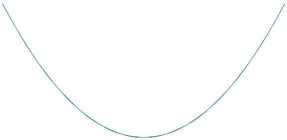
1 off-diagonal value (RS)



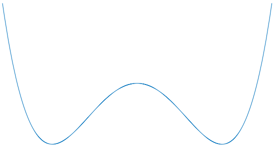
2 different off-diagonal values (1RSB)

RS corresponds to an asymptotically convex problem.

The Asymptotics of the Energy Function



1 off-diagonal value (RS)

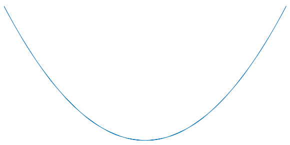


2 different off-diagonal values (1RSB)

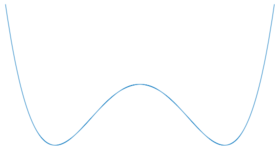
RS corresponds to an asymptotically convex problem.

For RSB, the decoupled channel cannot be Gaussian,

The Asymptotics of the Energy Function



1 off-diagonal value (RS)

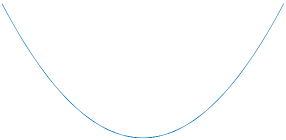


2 different off-diagonal values (1RSB)

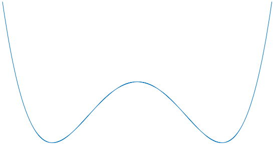
RS corresponds to an asymptotically convex problem.

For RSB, the decoupled channel cannot be Gaussian,
the noise must depend on the data,

The Asymptotics of the Energy Function



1 off-diagonal value (RS)

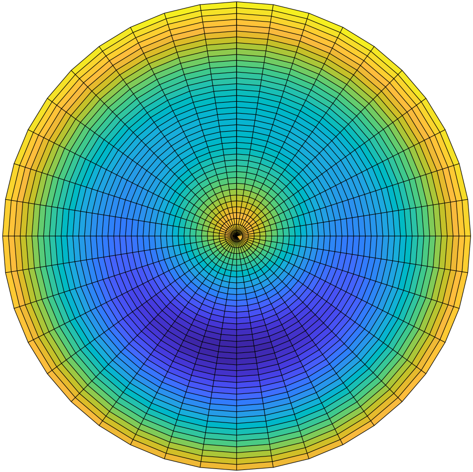


2 different off-diagonal values (1RSB)

RS corresponds to an asymptotically convex problem.

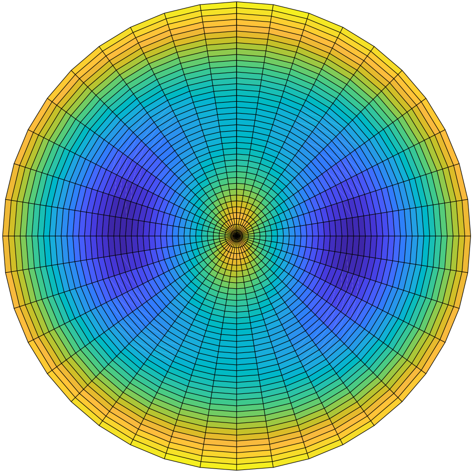
For RSB, the decoupled channel cannot be Gaussian,
the noise must depend on the data,
there can be more than 2 different values.

Asymptotic Energy Function vs. Number of Breaking Steps



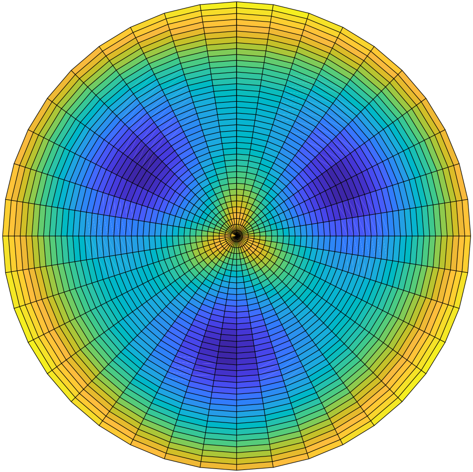
replica symmetric

Asymptotic Energy Function vs. Number of Breaking Steps



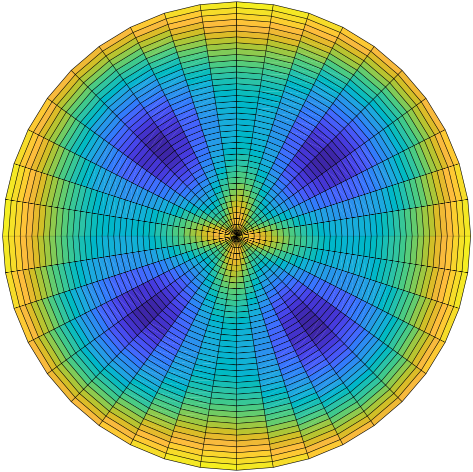
1 step of breaking

Asymptotic Energy Function vs. Number of Breaking Steps



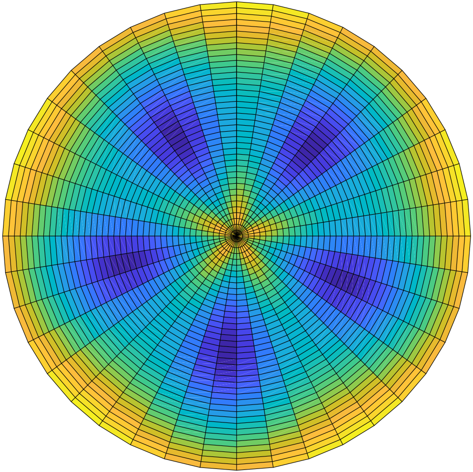
1 step of breaking

Asymptotic Energy Function vs. Number of Breaking Steps



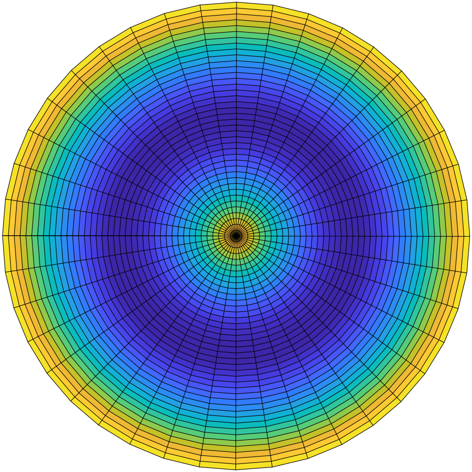
2 steps of breaking

Asymptotic Energy Function vs. Number of Breaking Steps



2 steps of breaking

Asymptotic Energy Function vs. Number of Breaking Steps



∞ steps of breaking

Entropy at Zero Temperature

… is a way to check whether a replica ansatz is too simplistic.

Let

$$\chi = \frac{\partial q_{kk}}{\partial T}.$$

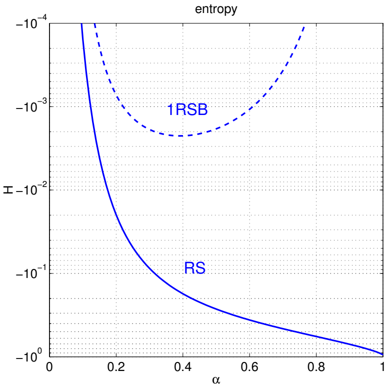
For quadratic energy function:

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

(BMS 2016)

Since $R(\cdot)$ is strictly increasing,

$$\chi > 0 \Leftrightarrow H < 0.$$



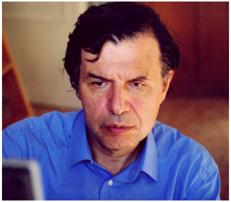
The closer the entropy is to zero, the better the RSB approximation.

A Historical Note

Replica symmetry breaking was introduced and solved by the replica method for a semicircularly distributed random matrix by Parisi in 1980 up to full RSB, i.e. infinite number of breaking steps.

His results were proven correct by rigorous means about 20 years later.

Nobel prize in physics in 2021.



Giorgio Parisi
born in Rome in 1948

Literature I

- F. Hiai, D. Petz. The Semicircle Law, Free Random Variables & Entropy, AMS '00.
- J. Mingo, R. Speicher. Free Probability and Random Matrix Theory, Springer '17.
- H. Nishimori. Statistical Physics of Spin Glasses and Information Processing, Oxford UP '01.
- M. Mezard, A. Montanari. Information, Physics & Computation, Oxford UP '09.
- A. Bereyhi, RM, H. Schulz-Baldes. Statistical mechanics of MAP estimation: General replica ansatz. IEEE Trans. Inform. Theory, vol. 65(12), 2019.

Free Energy for a Random Parameter

Consider a self-averaging random parameter, e.g. an (asymptotically large) measurement matrix $\boldsymbol{H}$.

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

The energy function depends on the random parameter $\boldsymbol{H}$, e.g. $||\boldsymbol{x}|| = \boldsymbol{x}^\dagger \boldsymbol{H}^\dagger \boldsymbol{H} \boldsymbol{x}$.

The expectation of a logarithm is a hard problem.

Replica Continuity

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

Evalute the n-th moments for integer n and assume analytic continuity for the limit.

More general, we have

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

Robert Brout introduced replicas in 1959 and first applied them to spin glasses in 1963.

Robert Brout born in 1928 in New York

Replica Continuity (cont'd)

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

With

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

we finally get

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

Replica Symmetry

Throughout the calculations, we solve integrals of the form

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

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

If the maximization is too tedious, we eventually 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.

Phase Transitions

If the final equations allow for multiple solutions, the correct solution is identified by

- minimizing the free energy, if exhaustive search is used to find the minimum,
- the worst solution, if a smooth descent algorithm is used to find the minimum.

While the free energy is smooth, other macroscopic functions may be discontinuous.

Roadmap for Large-System Analysis and Design

- Check if replica symmetry holds.
- Find the decoupled scalar channel.
- Analyze the decoupled scalar channel.
- Utilize the findings to implement the estimation of the data vector.

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

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

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

Zero Temperature Formulation

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

The quadratic form is written as a zero temperature limit

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

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

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

Replica Continuity

We want

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

Replica Continuity

We want

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

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

Replica Continuity

We want

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

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

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

Replica Continuity

We want

$$\lim_{K\to\infty}\frac{1}{K}\mathop{\mathrm{E}}_{\boldsymbol{J}}\log\sum_{\boldsymbol{x}\in\mathcal{X}}\mathrm{e}^{-\beta\mathrm{tr}(\boldsymbol{J}\boldsymbol{x}\boldsymbol{x}^\dagger)} = \lim_{K\to\infty}\lim_{n\to0}\frac{1}{nK}\log\mathop{\mathrm{E}}_{\boldsymbol{J}}\left(\sum_{\boldsymbol{x}\in\mathcal{X}}\mathrm{e}^{-\beta\mathrm{tr}(\boldsymbol{J}\boldsymbol{x}\boldsymbol{x}^\dagger)}\right)^n$$

$$= \lim_{K\to\infty}\lim_{n\to0}\frac{1}{nK}\log\mathop{\mathrm{E}}_{\boldsymbol{J}}\prod_{a=1}^{n}\sum_{\boldsymbol{x}_a\in\mathcal{X}}\mathrm{e}^{-\beta\mathrm{tr}\left(\boldsymbol{J}\boldsymbol{x}_a\boldsymbol{x}_a^\dagger\right)}$$

$$= \lim_{K\to\infty}\lim_{n\to0}\frac{1}{nK}\log\mathop{\mathrm{E}}_{\boldsymbol{J}}\sum_{\boldsymbol{x}_1\in\mathcal{X}}\cdots\sum_{\boldsymbol{x}_n\in\mathcal{X}}\mathrm{e}^{-\mathrm{tr}\left(\boldsymbol{J}\beta\sum_{a=1}^{n}\boldsymbol{x}_a\boldsymbol{x}_a^\dagger\right)}$$

$$= \lim_{K\to\infty}\lim_{n\to0}\frac{1}{nK}\log\mathop{\mathrm{E}}_{\boldsymbol{J}}\sum_{\boldsymbol{X}\in\mathcal{X}^n}\mathrm{e}^{-\mathrm{tr}\left(\boldsymbol{J}\beta\boldsymbol{X}\boldsymbol{X}^\dagger\right)}$$

Replica Continuity

We want

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

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

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

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

$$= \lim_{K\to\infty}\lim_{n\to 0}\frac{1}{nK}\log\sum_{\boldsymbol{X}\in\mathcal{X}^n}\exp\left[-K\sum_{a=1}^{n}\int\limits_0^{\lambda_a(\beta\boldsymbol{X}\boldsymbol{X}^\dagger/K)} R_{\boldsymbol{J}}(-w)\,\mathrm{d}w\right]$$

Just a Change of Variables

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

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

and find

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

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

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

Exponentials are Best

Write the Dirac measure as its inverse Laplace transform

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

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

Exponentials are Best

Write the Dirac measure as its inverse Laplace transform

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

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

In $n \times n$ dimensions, we have

$$\delta(\boldsymbol{Q}) = \oint_{\mathcal{C}^{n\times n}} \mathrm{e}^{\,\mathrm{tr}(\boldsymbol{QS})} \, \frac{\mathrm{d}\boldsymbol{S}}{(2\pi \mathrm{j})^{n^2}}.$$

Why Exponentials are Best

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

Why Exponentials are Best

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

Why Exponentials are Best

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

Now, the problem is no longer exponential in K.

Replica Symmetry

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

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

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

with some macroscopic parameters q and χ.

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

Twice Holds Better

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

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

We use f^2 for the parametrization, since later on we will use the Hubbard-Stratonovich transform, which will lead to square roots of S_{ab}.

Eigenvalues of $\boldsymbol{Q}$

For the evaluation of $G(\boldsymbol{Q})$, we can use replica symmetry to explicitly calculate the eigenvalues $\lambda_k(\boldsymbol{Q})$.
Consider the decomposition

$$\boldsymbol{Q} = \frac{\chi}{\beta}\mathbf{I} + q\mathbf{1}\mathbf{1}^\dagger.$$

Thus, the eigenvalues χ/β and $\chi/\beta + nq$ occur with multiplicities $n-1$ and 1, respectively. We get

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

Saddle Point Integration

Due to saddle point integration, the derivatives of

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

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

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

Taking derivatives yields

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

and solving for e and f gives

$$e = R(-\chi)$$
$$f = \sqrt{\frac{R(-\chi) - R(-\chi-\beta nq)}{\beta n}} \quad \xrightarrow{n\to 0} \quad \sqrt{qR'(-\chi)}.$$

The Taming of the Shrew

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

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

The Taming of the Shrew

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

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

The *Hubbard-Stratonovich transform*

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

tames it to read

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

The Law of Large Numbers

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

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

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

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

Saddle Point Integration Twice More

Partial derivatives of

$$\log M(e,f) - \mathrm{tr}(\boldsymbol{SQ})$$

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

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

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

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

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

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

Finally

Collecting previous results, we find with replica continuity that

$$E = \lim_{\beta\to\infty} \frac{1}{\beta} \lim_{n\to 0} \frac{\partial}{\partial n} \left[(n-1) \int_0^{\chi} R(-w)\,\mathrm{d}w + \int_0^{\chi+\beta n q} R(-w)\,\mathrm{d}w - \log M(e,f) \right.$$
$$\left. + n(n-1) f^2 \beta^2 q + n(f^2\beta - e)(\chi + \beta q) \right].$$

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

$$E = q\left[\mathrm{R}_{\boldsymbol{J}}(-\chi) - \chi \mathrm{R}'_{\boldsymbol{J}}(-\chi)\right].$$

1-Step Replica Symmetry Breaking (1RSB)

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

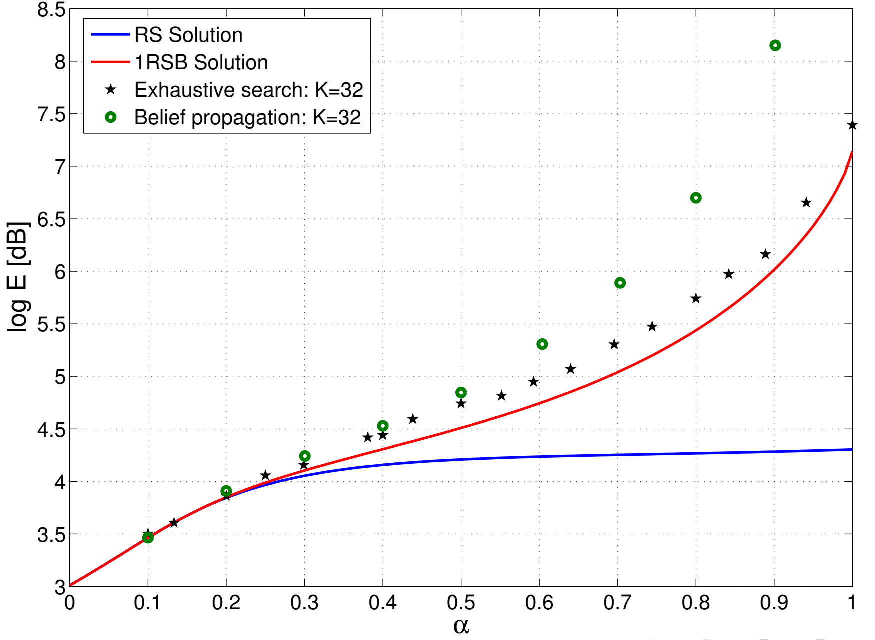
with the macroscopic parameters q, p and χ and the blocksize $\frac{\mu}{\beta}$.

1-Step Replica Symmetry Breaking

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

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

Solving those equations numerically is a tedious and tricky task.


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

Higher Step RSB Calculations

2RSB:

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 coupled fixed-point equations. Sometimes, they can be written as a functional equation.

Main Idea

Definition 1

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

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

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

Vector Notation

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

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

with $K > N$.

This is like overloaded CDMA.

Reconstruction by maximum a-posteriori detection

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

minimizes the probability that a wrong reconstruction is chosen.

Reconstruction Based on Minimum Distortion

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

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

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

Common distortion measures:

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

Common side constraints:

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

Average Distortion

Find the expected average distortion

$$\underset{\boldsymbol{x}|\boldsymbol{y}}{\mathrm{E}} \min_{\boldsymbol{\xi}} \mathrm{d}(\boldsymbol{x}, \boldsymbol{\xi}) = \iint \min_{\boldsymbol{\xi}} \mathrm{d}(\boldsymbol{x}, \boldsymbol{\xi}) \frac{\mathrm{p}_{\boldsymbol{y}|\boldsymbol{x}}(\boldsymbol{x}, \boldsymbol{y})\, \mathrm{p}_{\boldsymbol{x}}(\boldsymbol{x})}{\mathrm{p}_{\boldsymbol{y}}(\boldsymbol{y})} \mathrm{d}\boldsymbol{x} \mathrm{d}\boldsymbol{y}$$

- Write the minimum as a zero-temperature limit as in vector precoding.
- The conditional expectation is self-averaging as in CDMA.
- The self-average of the sensing matrix is along the lines of both vector precoding and CDMA.

Regularization

The marginalization over the prior is often np-hard. The approximation

$$\iint \min_{\boldsymbol{\xi}} \mathrm{d}(\boldsymbol{x}, \boldsymbol{\xi}) \frac{\mathrm{p}_{\boldsymbol{y}|\boldsymbol{x}}(\boldsymbol{x}, \boldsymbol{y})\, \mathrm{p}_{\boldsymbol{x}}(\boldsymbol{x})}{\mathrm{p}_{\boldsymbol{y}}(\boldsymbol{y})} \mathrm{d}\boldsymbol{x} \mathrm{d}\boldsymbol{y}$$

$$\approx \iint \min_{\boldsymbol{\xi}} \left[\mathrm{d}(\boldsymbol{x}, \boldsymbol{\xi}) + \mu f(\boldsymbol{\xi})\right] \frac{\mathrm{p}_{\boldsymbol{y}|\boldsymbol{x}}(\boldsymbol{x}, \boldsymbol{y})\, \check{\mathrm{p}}_{\boldsymbol{x}}(\boldsymbol{x})}{\mathrm{p}_{\boldsymbol{y}}(\boldsymbol{y})} \mathrm{d}\boldsymbol{x} \mathrm{d}\boldsymbol{y}$$

is called a regularization of the problem.

- Write the minimum as a zero-temperature limit as in vector precoding with the regularized objective function.
- The assumed conditional expectation is self-averaging as in CDMA.
- The self-average of the sensing matrix is along the lines of both vector precoding and CDMA.