Reconstructing Encrypted Signals:
Optimization with input from Spin Glasses and RMT.

Yan V Fyodorov

Department of Mathematics

KING'S College LONDON

Project supported by the EPSRC grant EP/N009436/1

Krakow, 3rd of May 2019

Content:

- **Part I:**
  Reconstructing nonlinearly encrypted signals corrupted by noise.
  YVF, *J. Stat. Phys.* (2019) [https://link.springer.com/article/10.1007/s10955-018-02217-9]

- **Part II:**
  On the loss function landscape in the simplest constrained least-square optimization
  Based on: **YVF**, R. Tublin under preparation

Rashel Tublin

Part I. Signal Reconstruction:

Background Model and Setting of the Problem:

Signals are represented by $N-$dimensional source (column) vectors $\mathbf{s} \in \mathbb{R}^N$. The associated **signal strength** R is defined via the Euclidean norm as

$$R = \sqrt{\frac{1}{N}(\mathbf{s}, \mathbf{s})}.$$

By a (symmetric key) **encryption** of the source signal we understand a **random mapping** $\mathbf{s} \mapsto \mathbf{y} \in \mathbb{R}^M$ known both to the sender and a recipient:

$$y_k = V_k(\mathbf{s}), \quad k = 1, \ldots, M \,,$$

where the collection of **random functions** $V_1(\mathbf{s}), \ldots, V_M(\mathbf{s})$ represents an encryption algorithm shared between the parties participating in the signal exchange.

Due to **imperfect** communication channels the recipients however get access to the encrypted signals only in a **corrupted form** modified by an additive random **noise**, i.e. $\mathbf{z} = \mathbf{y} + \mathbf{b}$ with the noise assumed to be normally distributed: $\mathbf{b} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{1}_M)$. A natural parameter is then the 'bare' **noise-to-signal** ratio (NSR) $\gamma = \sigma^2/R^2$.

The recipient's aim is to reconstruct the source signal $\mathbf{s}$ from the knowledge of $\mathbf{z}$.

Background Model and Setting of the Problem II:

We consider the reconstruction problem under a few technical assumptions:

- The recipient is aware of the exact source signal strength $R = \sqrt{\frac{1}{N}(\mathbf{s}, \mathbf{s})}$, and therefore can restrict the signal search to the feasibility set $\mathbb{W}$ given by $(N-1)-$dimensional sphere of the radius $R\sqrt{N}$.

- The random functions $V_k(\mathbf{s})$ belong to the class of (smooth) *isotropic* mean-zero Gaussian-distributed random fields on the sphere with the covariance structure dependent only on the angle between the vectors:

$$\langle V_k(\mathbf{x}) V_l(\mathbf{s}) \rangle = \delta_{lk} \Phi\left(\frac{(\mathbf{x}, \mathbf{s})}{N}\right) ,$$

  where the angular brackets $\langle \ldots \rangle$ denote the expected values. As our basic example we will consider the **linear-quadratic** family:

$$V_k(\mathbf{x}) = (\mathbf{a}_k, \mathbf{x}) + \frac{1}{2}(\mathbf{x}, \mathcal{J}^{(k)} \mathbf{x}) ,$$

  where $\mathbf{a}_k \sim \mathcal{N}(\mathbf{0}, \frac{J_1^2}{N} \mathbf{1}_N)$, and the entries of $N \times N$ real symmetric GOE-like random matrices $\mathcal{J}^{(k)}, k = 1, \ldots, M$ are mean-zero i.i.d. normal with the variance $\frac{J_2^2}{N^2}$. This results in the covariance of the form $\Phi(u) = J_1^2 u + \frac{1}{2} J_2^2 u^2$.

Background Model and Setting of the Problem III:

- We consider the input signal **s** through the reconstruction procedure as a *fixed* vector, and then employ the **Least-Square** reconstruction scheme, which for a given set of observations $z_k = V_k(\mathbf{s}) + b_k$ returns an estimate of the input signal as:

$$\mathbf{x} := Argmin_{\mathbf{w}} \left[\sum_{k=1}^{M} \frac{(z_k - V_k(\mathbf{w}))^2}{2} \right], \quad \mathbf{w} \in \mathbb{W} \subseteq \mathbb{R}^N ,$$

  where $\mathbb{W}$ is the sphere of feasible input signals. This scheme has the meaning of the Maximum–A-Posteriori (**MAP**) estimator with a uniform prior distribution over the sphere $\mathbb{W}$.

- The quality of the reconstruction will be characterized via the ratio

$$p_N := \frac{(\mathbf{x}, \mathbf{s})}{N R^2} \in [0, 1] ,$$

  where $p_N = 1$ corresponds to a reconstruction without any macroscopic distortion, whereas $p_N = 0$ manifests impossibility to recover any information from the originally encrypted signal.
  Our goal: Evaluate p_N for $N \gg 1$ as a function of the Noise-to-Signal ratio for a given degree of **redundancy** $\mu = M/N > 1$ and **nonlinearity** $a = R^2 J_2^2 / J_1^2$.

Remarks on the Method I:

Given the **fixed signal** **s** we interpret the **cost/loss** function

$$\mathcal{H}_{\mathbf{s}}(\mathbf{x}) = \sum_{k=1}^{M} \frac{(b_k + V_k(\mathbf{s}) - V_k(\mathbf{x}))^2}{2},$$

as an **energy** associated with a vector of N 'soft spins' $\mathrm{x}^T = (x_1, \ldots, x_N)$, with the configurations constrained to the sphere $\mathbb{W}$ of radius $|\mathbf{x}| = N\sqrt{R}$. In this way we can put the **least square** minimization problem in the context of **spin glass**-like Statistical Mechanics after introducing the inverse temperature parameter $\beta > 0$, and defining the partition function of the model as

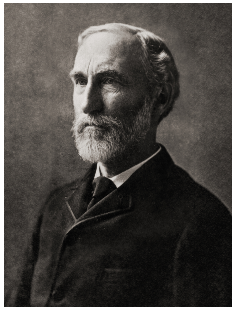
$$\mathcal{Z}_\beta = \int_{\mathbb{W}} e^{-\beta\mathcal{H}_{\mathrm{s}}(\mathbf{x})} d\mathbf{x}, \quad d\mathbf{x} = \prod_{i=1}^{N} dx_i\,.$$

We then consider the (Boltzmann) **Gibbs weights** $\pi_\beta(\mathbf{x}) = \mathcal{Z}_\beta^{-1} e^{-\beta\mathcal{H}_{\mathbf{s}}(\mathbf{x})}$ associated with any configuration $\mathbf{x}$ on the sphere $\mathbb{W}$. In the **zero-temperature** limit $\beta \to \infty$ the weights $\pi_\beta(\mathbf{x})$ concentrate on the set of globally minimal values of the cost function.

In particular, by considering

$$\left\langle p_N^{(\beta)} \right\rangle := \left\langle \frac{1}{\mathcal{Z}_\beta} \int_{\mathbb{W}} \frac{(\mathbf{x},\mathbf{s})}{NR^2}\, e^{-\beta\mathcal{H}_{\mathbf{s}}(\mathbf{x})} d\mathbf{x} \right\rangle_{V,\mathbf{b}}$$

we aim to evaluating $p_\infty := \lim_{\beta\to\infty} \lim_{N\to\infty} \left\langle p_N^{(\beta)} \right\rangle$ providing us with a measure of the quality of the signal reconstruction.



J W Gibbs (1839–1903)

Main Results for General Nonlinearity I:

Given the source signal strength $R > 0$, and the redundancy $\mu = M/N > 1$, the **mean value** of the parameter p_N characterising quality of the information recovery in the **Least-Square** reconstruction scheme with the noise $\mathbf{b} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{1}_M)$ is given asymptotically for $N \to \infty$ by

$$p_\infty := \lim_{N\to\infty} \langle p_N \rangle = \tfrac{t}{R} \,,$$

where the specific value of $t \in [0, R]$ should be found in the framework of the **Parisi** scheme of the **Full Replica Symmetry Breaking** (FRSB) by **minimizing** the functional

$$\mathcal{E}[w_s(u); Q, v, t] = -\left[\frac{R^2 - t^2 - Q}{v + \int_{R^2-Q}^{R^2} w_s(u)\, du} + \int_{R^2-Q}^{R^2} \frac{dq}{v + \int_q^{R^2} w_s(u)\, du}\right]$$
$$+\mu\left[\frac{\sigma^2 + \Phi(R^2) - 2\Phi(Rt) + \Phi(R^2 - Q)}{1 + v\Phi'(R^2) + \int_{R^2-Q}^{R^2} w_s(u)\Phi'(u)\, du} + \int_{R^2-Q}^{R^2} \frac{\Phi'(q)\, dq}{1 + v\Phi'(R^2) + \int_q^{R^2} w_s(u)\Phi'(u)\, du}\right],$$

over t, and **maximizing** it over all the variables $v \geq 0$ and $Q \in [0, R^2]$ and over a non-decreasing function $w_s(u)$ with the argument $u \in [R^2 - Q, R^2]$.

Giorgio Parisi

Main Result for General Nonlinearity II:

- In a certain range of parameters (e.g. the redundancy and nonlinearity) the above variational problem is solved by the **Replica-Symmetric** Ansatz $Q = 0$. In that case for a given 'bare' Noise-to-Signal ratio $\gamma = \sigma^2/R^2$ the quality parameter $p_\infty = p \in [0,1]$ is given by the solution of a **single** algebraic equation:

$$p^2\left(\gamma + 2\,\frac{\Phi(R^2)-\Phi(R^2p)}{R^2}\right) = \mu(1-p^2)\frac{\left[\Phi'(R^2p)\right]^2}{\Phi'(R^2)}\,.$$

- For the alternative range of parameters the variational problem can be solved by the **FRSB Ansatz** assuming the minimizer function $w_s(u)$ to be **continuous** and **non-decreasing** for $u \in [R^2-Q, R^2]$. In that case the value $p_\infty = p$ is given by the solution of the system of a **pair** of algebraic equations in the variables $p \in [0,1]$ and $Q \in (0, R^2]$:

$$\begin{aligned}&\mu\,\left[\Phi'(R^2p)\right]^2\left(R^2(1-p^2)-Q)\right)\\ &= p^2\Phi'(R^2-Q)\left[R^2\gamma+\Phi(R^2)-2\Phi(R^2p)+\Phi(R^2-Q)\right]\end{aligned}$$

and

$$\left[\Phi'(R^2-Q)\right]^3 p^2 = \mu\left[\Phi'(R^2p)\right]^2\left[\Phi'(R^2-Q)-\Phi''(R^2-Q)\left(R^2(1-p^2)-Q\right)\right]$$

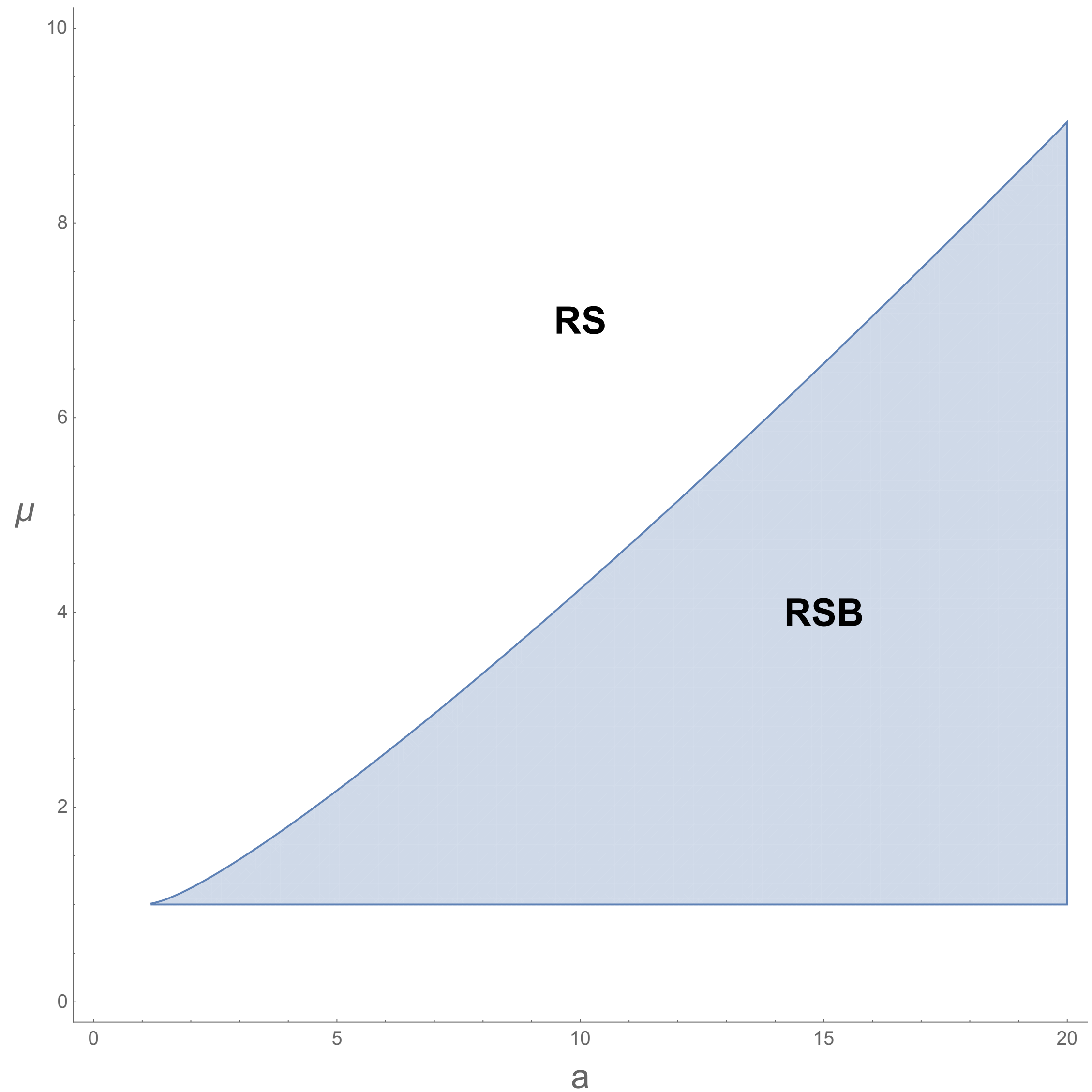


Figure 1: Schematic Phase diagram in $(a = J_2^2/J_1^2, \mu = M/N)$ plane for **Linear-Quadratic** encryptions. In the shaded region of parameters $1 < \mu < \frac{(a^{2/3} - a^{1/3} + 1)^3}{a}$ replica symmetry must be fully broken for some amplitude of the noise.

Reconstruction quality for purely quadratic encryptions a $= \infty$:

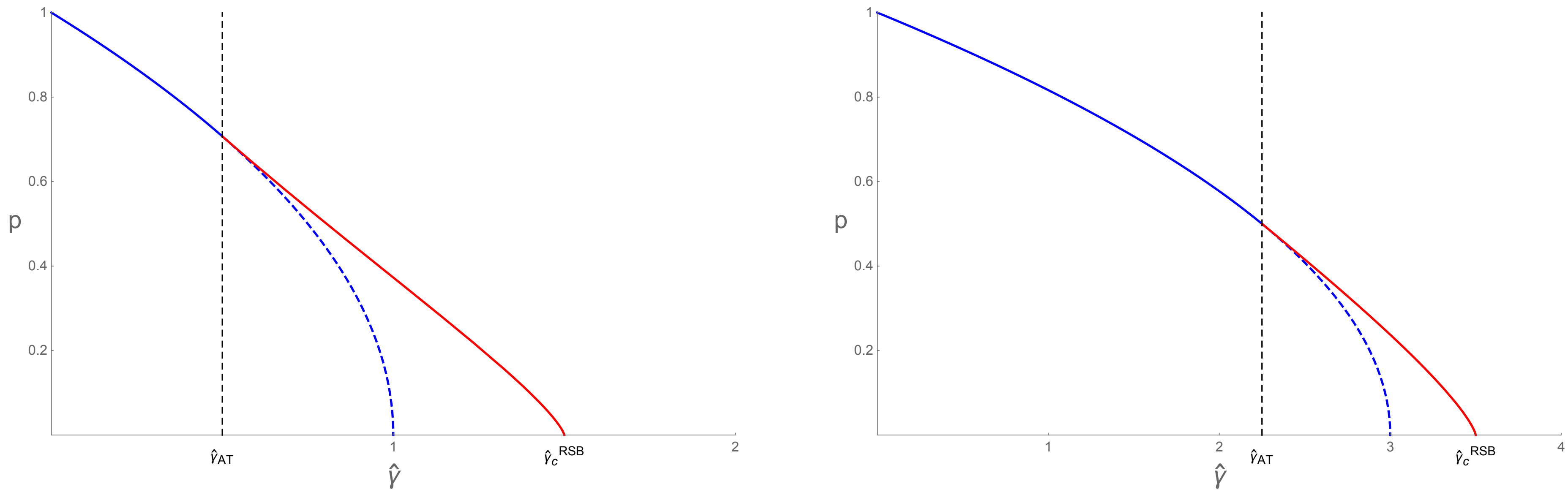


Figure 2: The **quality parameter** p as a function of the scaled **noise-to-signal** ratio $\hat{\gamma} = \frac{\sigma^2}{J_2^2 R^4}$ for **purely quadratic** encryptions and two different redundancies: $\mu = 2$ (left) and $\mu = 4$ (right). There always exists a **threshold value** $\hat{\gamma}_c(\mu)$ such that $p_\infty = 0$ for $\hat{\gamma} > \hat{\gamma}_c(\mu)$ making the reconstruction **impossible** beyond some level of noise. The behaviour close to the threshold is given by $p_\infty \sim (\hat{\gamma}_c - \hat{\gamma})^{3/4}$ and is controlled by the **replica symmetry breaking** mechanism.
The blue broken curve is the continuation of the replica-symmetric solution in the region of Full RSB.

Open questions:

The problem is shown to be equivalent to finding the configuration of minimal energy in a certain version of spherical spin glass model, with **squared** Gaussian random interaction potential. It would be interesting and instructive, in particular,

- to develop **rigorous** approach to this type of landscapes beyond replicas, in particular to study **complexity** associated with the stationary points/minima. So far we managed to do it only for the special type of **purely linear** Least Square schemes (with **R. Tublin**, part II)

- to study fluctuations in the overlap and/or in the depth of global minimum, etc.

- Analyze gradient search dynamics on the sphere.

Part II. Loss function Landscape in the simplest case:

The simplest optimization problem of the **least-square** type on the sphere $\mathbf{x} \in \mathbb{R}^N, \ \mathbf{x}^2 = const$ arises in the **Multiple Factor Data Analysis** and is known as the **Oblique Procrustes Problem**:

For a given pair of $M \times N$ matrices $\mathbf{A}$ and $\mathbf{B}$ find such $N \times N$ matrix $\mathbf{X}$ that the equality $\mathbf{B} = \mathbf{AX}$ holds as close as possible and columns $\mathbf{x}_i \in \mathbb{R}^N$, $i = 1, \ldots N$ are of unit length.

For $M > N$ this system of linear equations is overcomplete and a solution can be found separately for each column $\mathbf{x}$ by minimizing the **loss/cost function**

$$H(\mathbf{x}) = \frac{1}{2}||A\mathbf{x} - \mathbf{b}||^2 := \frac{1}{2}\sum_{k=1}^{M}\left[\sum_{j=1}^{N} A_{kj}\mathbf{x}_j - b_k\right]^2, \quad \mathbf{x}^2 = const$$

The problem was first analysed in that setting by **M. W. BROWNE** in 1967, and then independently by numerical mathematicians (e.g. **W. GANDER** 1981) who used the **Lagrange multiplier** to take care of the spherical constraint. Introducing the Lagrangian $\mathcal{L}_{\lambda,\mathrm{s}}(\mathbf{x}) = \mathcal{H}(\mathbf{x}) - \frac{\lambda}{2}(\mathbf{x}, \mathbf{x})$, with real λ being the Lagrange multiplier, the stationary conditions $\nabla\mathcal{L}_{\lambda,\mathrm{s}}(\mathbf{x}) = 0$ yields linear system:

$$A^T\left[A\mathbf{x} - \mathbf{b}\right] = \lambda\mathbf{x}, \quad \Rightarrow \mathbf{x} = (A^TA - \lambda I_N)^{-1}A^T\mathbf{b}$$

Setting of the problem:

The spherical constraint $\mathbf{x}^2 = N$ yields the equation for λ in the form:

$$\mathbf{b}^T A \, \frac{1}{\left(A^T A - \lambda I_N\right)^2} \, A^T \mathbf{b} = N$$

which is equivalent to a polynomial equation of degree $2N$ in λ. Each **real** solution for the **Lagrange multiplier** λ_i corresponds to a **stationary point** $\mathbf{x}_i$ of the loss function $H(\mathbf{x}) = \frac{1}{2}||A\mathbf{x} - \mathbf{b}||^2$ on the sphere $\mathbf{x}^2 = N$ and one can show that the order $\lambda_1 < \lambda_2 < \ldots < \lambda_{\mathcal{N}}$ implies $H(\mathbf{x}_1) < H(\mathbf{x}_j) < \ldots < H\left(\mathbf{x}_{\mathcal{N}}\right)$. Thus the **minimal loss** is given by $\mathcal{E}_{min} = H\left(\mathbf{x}_1\right)$.

Our goal: To count the **stationary points** via the Lagrange multipliers

$$\lambda_i, \; i = 1, \ldots, \mathcal{N} \leq 2N$$

and eventually find the **minimal loss** $\mathcal{E}_{min}$ after assuming the entries A_{kj} of $M \times N$, $M > N$ matrix A to be i.i.d. normal real variables such that $A^T A = W$ is $N \times N$ **Wishart** matrix with the probability density

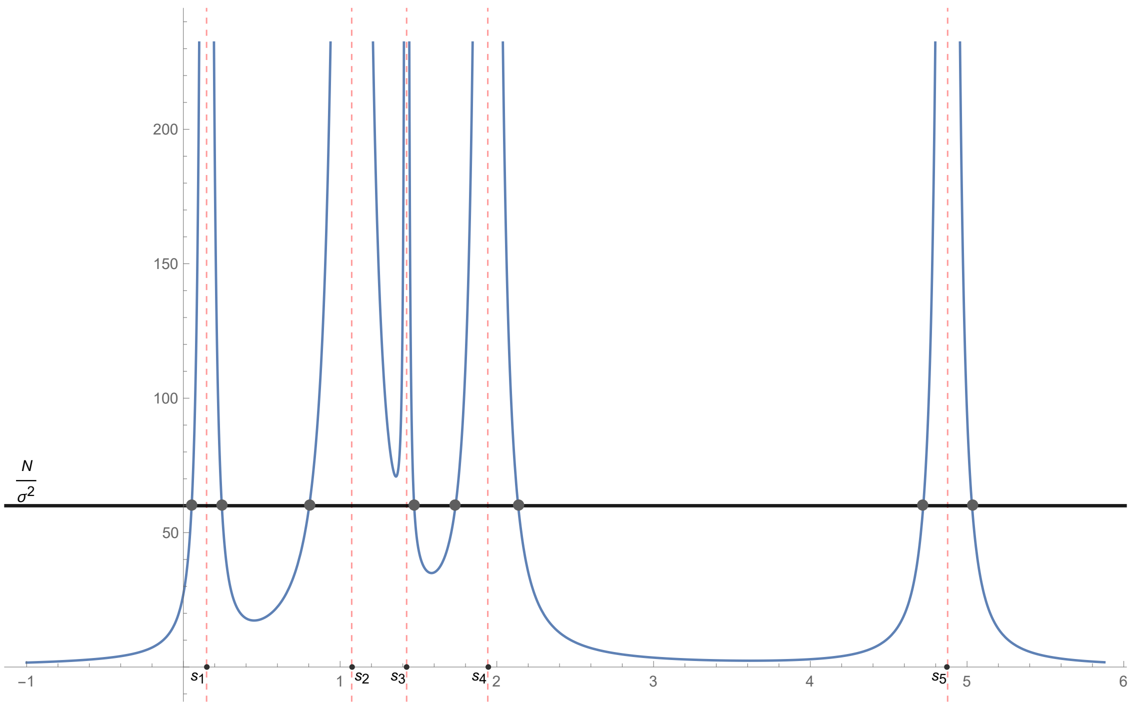
$$P_{N,M}(W) = C_{N,M} e^{-\frac{N}{2}\mathrm{Tr} W} \left(\det W\right)^{\frac{M-N-1}{2}}$$

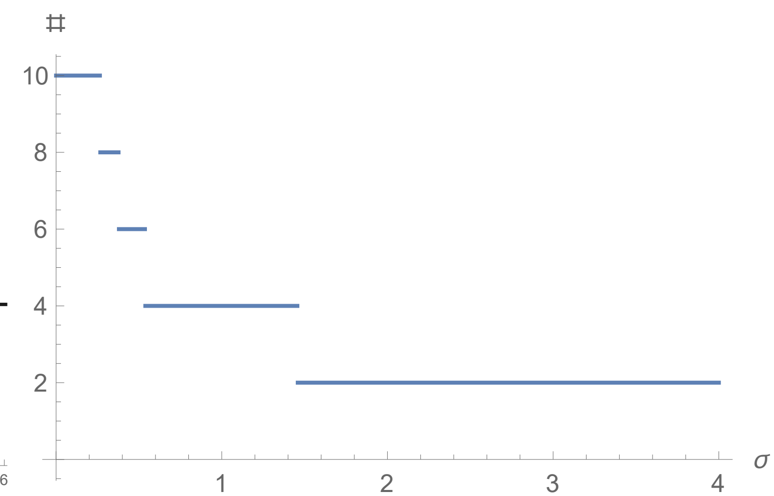
We will also assume for convenience that the vector $\mathbf{b}$ *is normally distributed:* $\mathbf{b} = \sigma \, \xi$ *with* $\sigma > 0$ *and the components of* $\xi = (\xi_1, \ldots, \xi_M)^T$ *are mean zero standard normals.*

Qualitative considerations:

The equation for the Lagrange multiplier can be conveniently written in terms of N nonzero eigenvalues $s_1, \ldots, s_N$ of $M \times M$ matrix $W^{(a)} = AA^T$ and the associated eigenvectors $\mathbf{v}_i$:

$$\sum_{i=1}^{N} \frac{s_i}{(\lambda - s_i)^2} (\xi^T \mathbf{v}_i)^2 = \frac{N}{\sigma^2}$$





Case $N = 5$

Counting zeroes via Kac-Rice formula:

Mark Kac (1914-1984) and **Stephen O. Rice** (1907-1986)

Number $\mathcal{N}_{(a,b)}$ of simple zeroes of a (smooth enough) function $f(x)$ in $x \in (a,b)$ can be found via

$$\mathcal{N}_{(a,b)} = \int_a^b \delta(f(x))|f'(x)|\,dx$$

Counting Lagrange multipliers via the Kac-Rice formula:

The number $\mathcal{N}_{st}[a,b]$ of real solutions of the equation $A^T\left[A\mathbf{x}-\mathbf{b}\right]-\lambda\mathbf{x}=0$ on the sphere $\mathbf{x}^2=N$ such that $\lambda\in[a,b]$ can be counted by employing the **Kac-Rice** type formula

$$\mathcal{N}_{st}[a,b]=\int_a^b d\lambda\int\delta\left[A^T\left(A\mathbf{x}-\mathbf{b}\right)-\lambda\mathbf{x}\right]\delta\left(\mathbf{x}^2-N\right)$$
$$\times\left|\det\begin{pmatrix}A^TA-\lambda I_N & \mathbf{x}\\ -2\mathbf{x}^T & 0\end{pmatrix}\right|d\mathbf{x}$$

Using Gaussianity of both the matrix entries $A_{ij}\sim\mathcal{N}(0,1)$ and the vector components $\mathbf{b}\sim\mathcal{N}_M(0,I_M\sigma^2)$ and introducing the parameter $\delta=\frac{1}{2}\ln\left(1+\sigma^2\right)$ one can eventually find the mean number of solutions as

$$\mathbb{E}\left\{\mathcal{N}_{st}[a,b]\right\}=\int_a^b p(\lambda)\,d\lambda$$

with the density $p(\lambda)$ for $\lambda>0$ given by

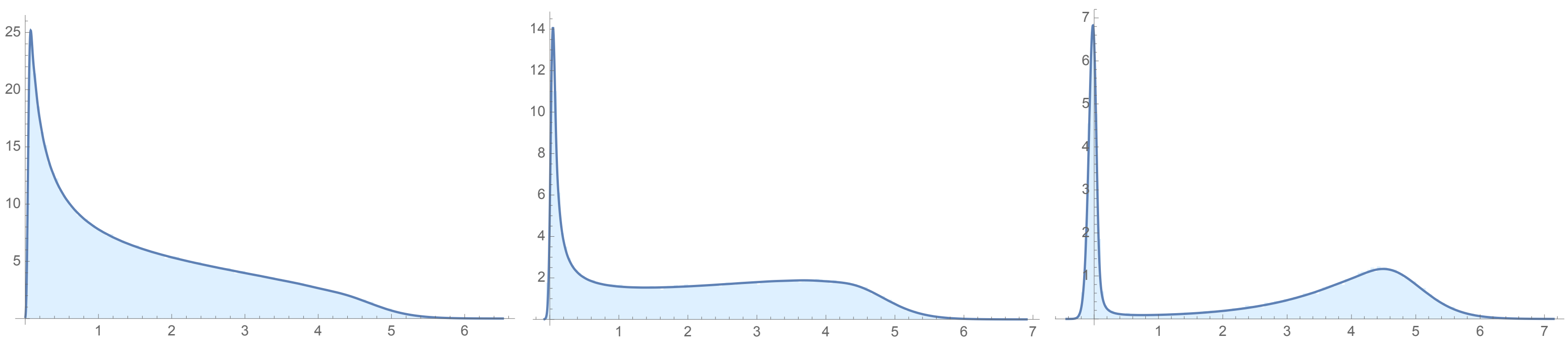
$$p(\lambda\geq 0)=2\sqrt{\frac{N}{\pi}}\frac{e^{-\frac{M+N-1}{2}\delta}}{\sqrt{\sinh\delta}}K_{\frac{M-N}{2}}\left(\frac{N\lambda}{2\sinh\delta}\right)e^{\frac{N\lambda}{2}\coth\delta}\left\langle\rho_N(\lambda)\right\rangle\sqrt{\lambda}$$

where $K_\nu(z)$ is the Bessel-Macdonald function, and $\langle\rho_N(\lambda)\rangle$ stands for the mean eigenvalue density of $N\times N$ Wishart matrix $W=A^TA$ presented for any M, N in **Introduction to Random Matrices: Theory and Practice** by **G. Livan**,**M. Novaes** and **P. Vivo** (Springer 2018).

Counting Lagrange multipliers via the Kac-Rice formula :

For negative values of the Lagrange multiplier λ we have instead:

$$p(\lambda<0)=\frac{N!N^{(M-N)/2}}{2^{(M+N-3)/2}}\frac{1}{\Gamma\left(\frac{N}{2}\right)\Gamma\left(\frac{M}{2}\right)}\frac{e^{-(M+N-1)\delta/2}}{\sqrt{\sinh\delta}}e^{-\frac{1}{2}N|\lambda|(\coth\delta-1)}|\lambda|^{(M-N)/2}$$
$$\times\left[\sum_{j=0}^{N-1}\binom{M-1}{N-1-j}\frac{1}{j!}(N|\lambda|)^{j}\right]K_{\frac{M-N}{2}}\left(\frac{N|\lambda|}{2\sinh\delta}\right)$$



Evolution of the density $p(\lambda)$ for $N=20\,,\ M=30$
as the function of variance $\sigma^2=0.005;0.25;0.70$

The blue histograms correspond to 10000 realizations.

Large Deviations for the smallest Lagrange multiplier:

For large $N \to \infty$, fixed $1 < \mu = M/N < \infty$ and fixed finite $\sigma^2 > 0$ the probability density for the smallest Lagrange multiplier λ_{min} has the **Large Deviation** form:

$$p(\lambda_{min})|_{\lambda_{min}<s_-} \sim e^{-\frac{N}{2}\boldsymbol{\Phi}(\lambda_{\min})}, \quad \boldsymbol{\Phi}(\lambda) = \mathbf{L_1}(\lambda) + \mathbf{L_2}(\lambda) + \frac{(\mu+1)}{2}\ln\left(1+\sigma^2\right),$$

where $s_- = (\sqrt{\mu}-1)^2$ is the '**Marchenko-Pastur**' left edge and for $\kappa = \frac{(\mu-1)\sigma^2}{2\sqrt{1+\sigma^2}}$

$$\mathbf{L_1}(\lambda) = (\mu-1)\left\{\frac{\sqrt{\lambda^2+\kappa^2}}{\kappa} - \ln\left(\kappa+\sqrt{\lambda^2+\kappa^2}\right) - \lambda\frac{\sqrt{(\mu-1)^2+\kappa^2}}{(\mu-1)\kappa}\right\}$$

$$\mathbf{L_2}(\lambda) = -\sqrt{(\lambda-s_-)(\lambda-s_+)} - 2\ln\frac{\left(\mu+1-\lambda+\sqrt{(\lambda-s_-)(\lambda-s_+)}\right)}{2\sqrt{\mu}}$$

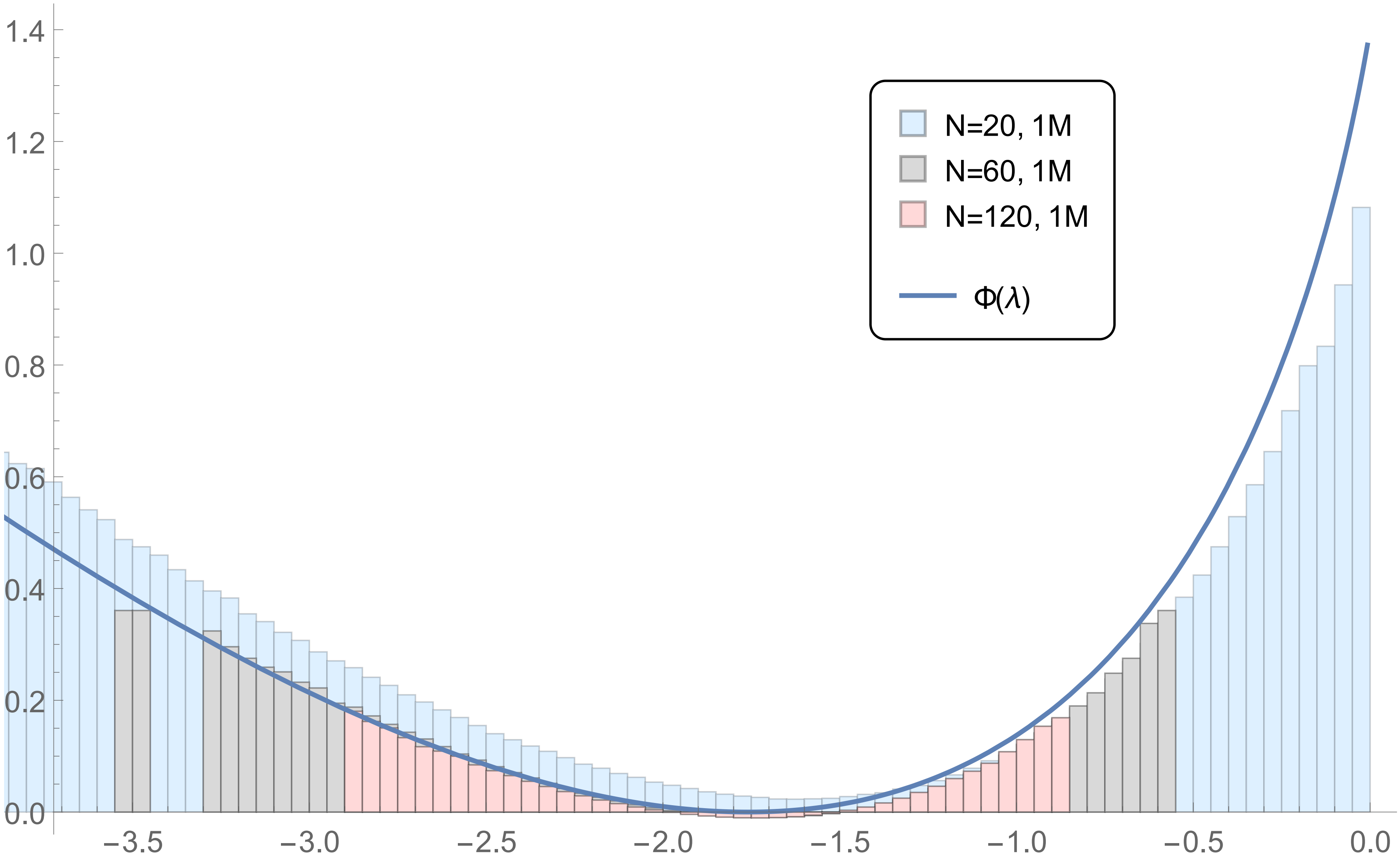
$$+2(\mu-1)\ln\frac{\left(\mu-1+\lambda+\sqrt{(\lambda-s_-)(\lambda-s_+)}\right)}{2\sqrt{\mu}}$$

One finds that $\Phi(\lambda)$ is **minimized** for

$$\lambda = \lambda_* = \left(\sqrt{\mu}-\sqrt{1+\sigma^2}\right)\left(\sqrt{\mu}-\frac{1}{\sqrt{1+\sigma^2}}\right)$$

which eventually implies the **most probable** value of the **minimal loss/error**:

$$\lim_{N\to\infty}\frac{\mathcal{E}_{min}}{N} = \frac{1}{2}\left[\sqrt{\mu(1+\sigma^2)}-1\right]^2$$



The large deviation function for the smallest Lagrange multiplier vs. simulations

Conclusions:

- We counted the mean number of **stationary points** of the simplest '**least-square**' optimization problem on a sphere via the Lagrange multipliers in various scaling regimes, and found the **typical** minimal loss $\mathcal{E}_{min}$.

- **Open questions:**
  - Fluctuations of the counting function,
  - **large/small deviations** of the minimal loss $\mathcal{E}_{min}$
  - Gradient search dynamics on the sphere
  - Landscape for a **nonlinear** 'least-square' optimization, etc.

Conclusions:

- We counted the mean number of **stationary points** of the simplest '**least-square**' optimization problem on a sphere via the Lagrange multipliers in various scaling regimes, and found the **typical** minimal loss $\mathcal{E}_{min}$.

- **Open questions:**

  – Fluctuations of the counting function,
  – **large/small deviations** of the minimal loss $\mathcal{E}_{min}$
  – Gradient search dynamics on the sphere
  – Landscape for a **nonlinear** 'least-square' optimization, etc.

THANK YOU!