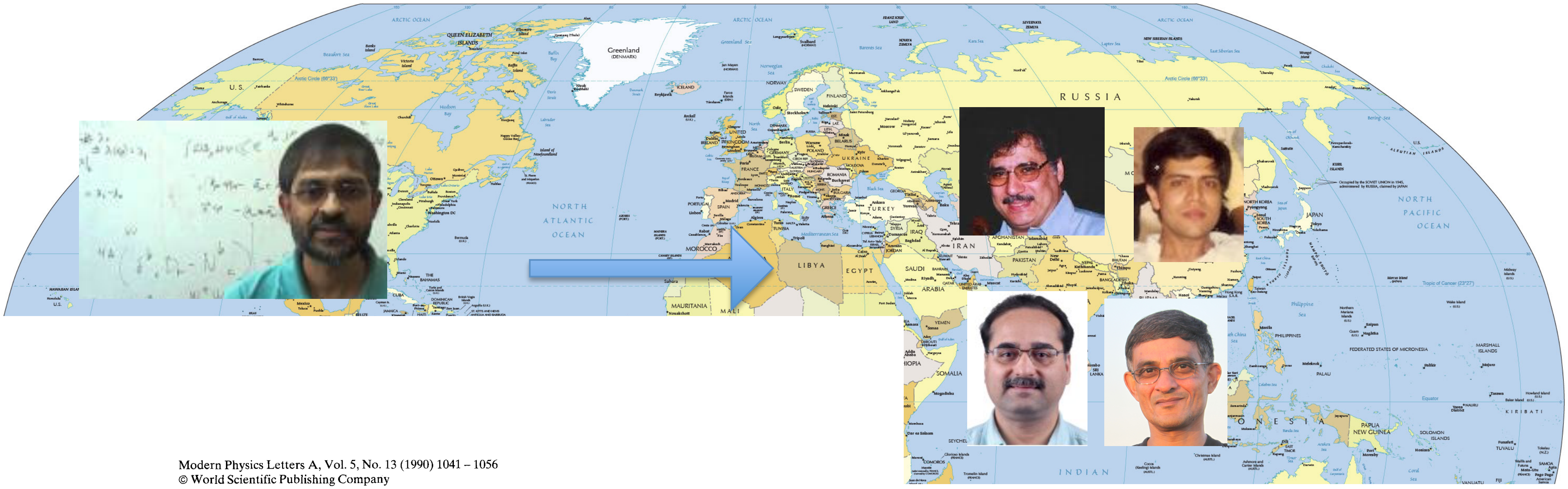
High-dimensional statistics and large-SpeNta expansion OR One can go only so far from home

Anirvan Sengupta
Physics and Astronomy
Rutgers University

TIFR String Excitations: 1987-88

??
R. Saroja

Fall 1989



Modern Physics Letters A, Vol. 5, No. 13 (1990) 1041 – 1056

NEW CRITICAL BEHAVIOR IN $d = 0$ LARGE-N MATRIX MODELS

SUMIT R. DAS, AVINASH DHAR, ANIRVAN M. SENGUPTA and SPENTA R. WADIA
Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400 005, India

Received 9 January 1990

The non-perturbative formulation of 2-dimensional quantum gravity in terms of the large-N limit of matrix models is studied to include the effects of higher order curvature terms. This leads to matrix models whose potential contains a symmetry breaking term of the form Tr $\phi A \phi A$, where A is a given fixed matrix. This is studied in $d = 0$ dimensions and effectively induces additional terms of the form $(\mathrm{Tr}\,\phi^k)^2$ in the one matrix potential. An exact solution to leading order of the potential $V(\phi) = 1/2\,\mathrm{Tr}\,\phi^2 + g/N\,\mathrm{Tr}\,\phi^4 + g'/N^2\,(\mathrm{Tr}\,\phi^2)^2$ is presented leading to 3 phases: $\gamma = -1/2$ (smooth surfaces), $\gamma = 1/2$ (branched polymer) and $\gamma = 1/3$ (intermediate phase). Including a Tr ϕ^6 term in the potential gives rise to an additional phase with $\gamma = 1/4$. It is conjectured that for the general polynomial potential there are phases with $\gamma = 1/n$, $n = 2, 3, \ldots$. The $\gamma > 0$ phases may correspond to $c > 1$ matter coupled to 2-dimensional gravity.

, *USA*

1. Introduction

It is generally believed that many fundamental issues in string theory are non-perturbative in nature. In this respect the standard first quantized approach seems to be of little use since it is always restricted to two-dimensional surfaces and hence to the perturbation expansion. Nevertheless, it has been realized recently that a deeper understanding of the integration over two-dimensional metrics in the first-

lidean quantum
osed surfaces of

Where was Spenta? (late eighties)

Spring 1991

Modern Physics Letters A, Vol. 6, No. 18 (1991) 1685–1692

CLASSICAL SOLUTIONS OF 2-DIMENSIONAL STRING THEORY

GAUTAM MANDAL*
School of Natural Sciences, Institute for Advanced Study, Princeton, NJ 08540, USA
ANIRVAN M. SENGUPTA
School of Natural Sciences, Institute for Advanced Study, Princeton, NJ 08540, USA
and Tata Institute of Fundamental Research, Bombay 400 005, India
and
SPENTRA R. WADIA*
School of Natural Sciences, Institute for Advanced Study, Princeton, NJ 08540, USA

Received 25 March 1991

We present an exact one-parameter family of solutions to the classical graviton-dilaton system in two dimensions. The solution can be identified as a black hole. We present the solution both in a Schwarzschild-like gauge and in the target space conformal gauge. We discuss possible relations with matrix models.

1. Introduction

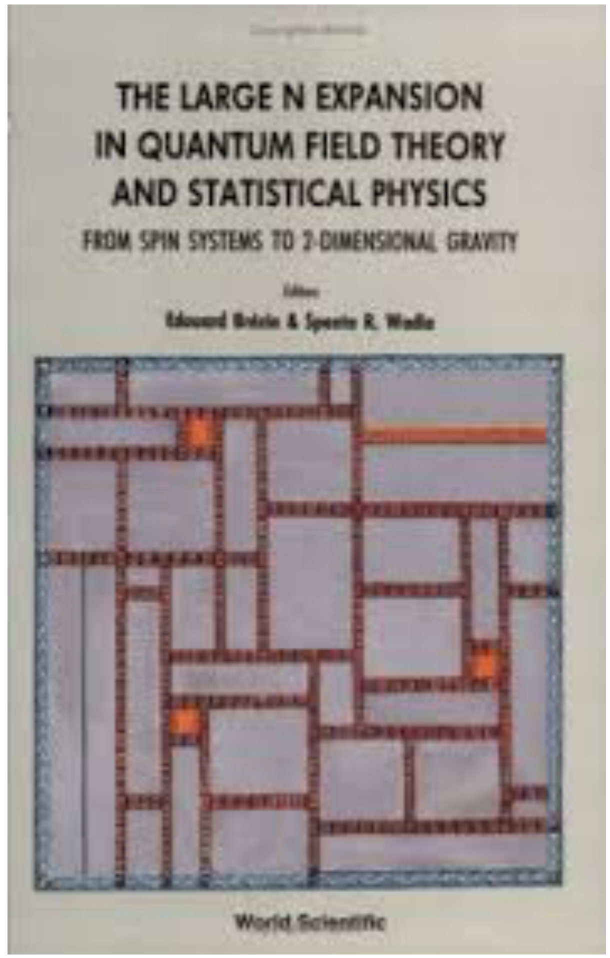
Recently there has been considerable interest in the problem of quantized 2-dimensional gravity coupled to a scalar field, which in the absence of gravity would be a $c = 1$ system. There are presently two approaches to this problem — the matrix model[1, 3–9] and the continuum Liouville theory.[10–12] There are encouraging results which indicate that the weak coupling tree level S-matrices computed in the matrix model can be obtained by Liouville theory calculation.[a] One of the interesting results of the continuum Liouville theory is that this system has a 2-dimensional target space, in that the general couplings are functions of two variables one of them being the conformal mode of the 2-dimensional metric.[2] The reasoning that led to this proposal was similar to that which leads to the σ-model approach to critical

*On leave from: Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400 005, India.

[a]By a Liouville theory here we mean a string in 2-dimensional target space with the following backgrounds: $G_{\mu,\nu} = \eta_{\mu\nu}$ (metric), $\Phi = Q\eta/2$ (dilaton) and $T = a\exp(Q\eta/2)$ (tachyon), where η is the Liouville mode. However, the 2-D gravity theory coupled to matter can admit of more general backgrounds having more complicated dependences on η. The reasoning that led to this proposal[2] is similar to the one that led to the σ-model approach in critical string theory.

ICTS_1

Large N Expansion



Recurring theme we heard today

- Emphasis on physical intuition
- Emphasis on unity of physics

Large N expansion as a fixed point

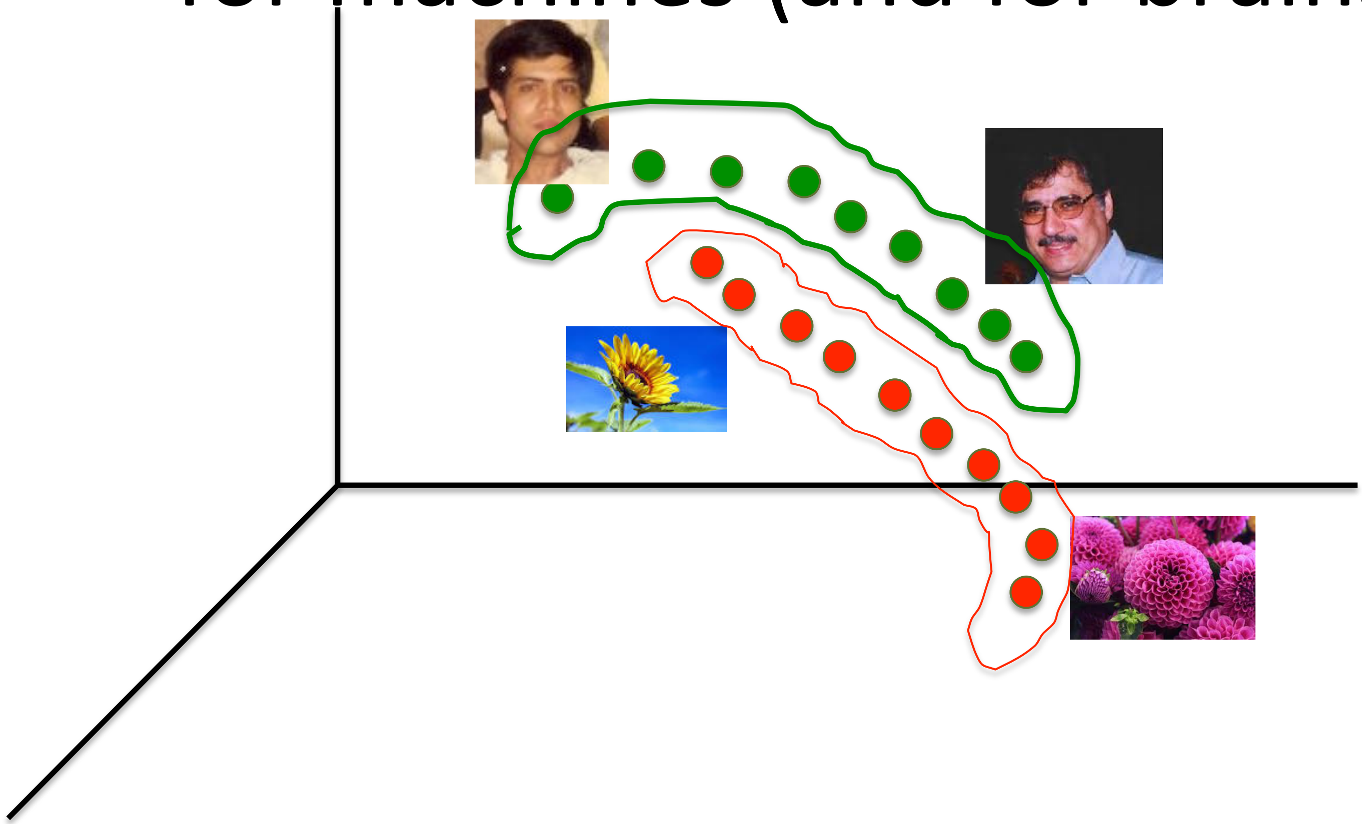
Kondo Physics:
Sengupta, Georges (1995)
Parcollet, Georges, Kotliar, Sengupta (1998)

Singular Value Distribution of Random Matrices:
Sengupta, Mitra (1999)

Information Theory of MIMO Systems:
Moustakas, Baranger, Balents, Sengupta, Simon (2000)
Moustakas, Simon, Sengupta (2003)

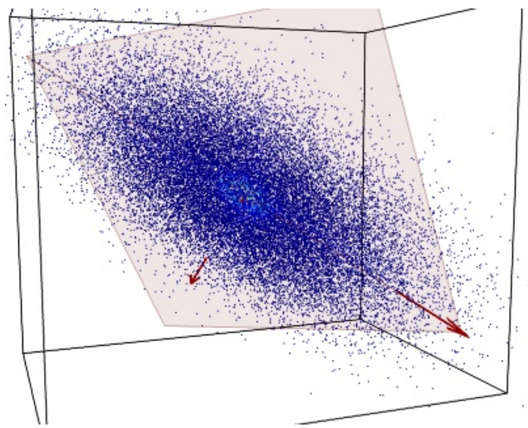
Algorithmic Phase Transitions in Compressed Sensing:
Ramezanali, Mitra, Sengupta (2015)

High dimensional pattern recognition: for machines (and for brains)



Curse of dimensionality

Dimension Reduction 101: Principal Components Analysis (PCA)



Diagonalize covariance matrix with Element given by:

$$C_{ij} = \sum_{a} (X_{ia} - \bar{X}_i)(X_{ja} - \bar{X}_j)$$

http://www.hgu.mrc.ac.uk/Research/Hill/Fig_4.jpg

PCA-> Singular Value Decomposition (SVD)

If we define $\hat{X}_{ia} = X_{ia} - \bar{X}_i$ then $C = \hat{X}\hat{X}^T$

Diagonalizing covariance matrix *C* is the same thing as doing a singular value decomposition of deviations *X* hat, and ordering singular vectors according to the size of singular values.

$$\underset{m \times n}{\overset{\hat{X}}{\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & \\ \vdots & \vdots & \ddots & \\ x_{m1} & & & x_{mn} \end{pmatrix}}} \approx \underset{m \times r}{\overset{U}{\begin{pmatrix} u_{11} & \cdots & u_{1r} \\ \vdots & \ddots & \\ u_{m1} & & u_{mr} \end{pmatrix}}} \underset{r \times r}{\overset{S}{\begin{pmatrix} s_{11} & 0 & \cdots \\ 0 & \ddots & \\ \vdots & & s_{rr} \end{pmatrix}}} \underset{r \times n}{\overset{V^{\mathsf{T}}}{\begin{pmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \\ v_{r1} & & v_{rn} \end{pmatrix}}}$$

SVD for Denoising/Compressing Images

Figure 2: 1^{st}, 50^{th}, and 120^{th} image approximations of the image for $\sigma = 0\%$

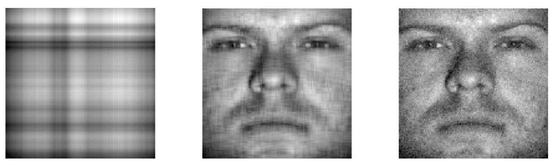


Figure 3: 1^{st}, 14^{th}, and 120^{th} image approximations of the image for $\sigma = 5\%$

Figure 4: 1^{st}, 9^{th}, and 120^{th} image approximations of the image for $\sigma = 10\%$

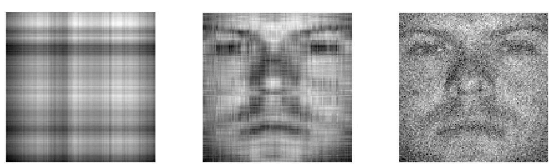


Figure 5: 1^{st}, 5^{th}, and 120^{th} image approximations of the image for $\sigma = 15\%$

From Poon, Ng and Sridharan, 2009

$$X_{\mathrm{recon}} = \bar{X} + U S_K V^{\dagger}$$

$$S_K = \begin{pmatrix} s_{11} & \cdots & \cdots & 0\cdots & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \cdots & \vdots \\ 0 & \cdots & s_{KK} & \vdots & \cdots & \vdots \\ 0 & \cdots & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & \cdots & 0 \end{pmatrix}$$

Q: What *K* is optimal?

A: Departure from singular values of pure noise.

Describe the distribution in a different coordinate system

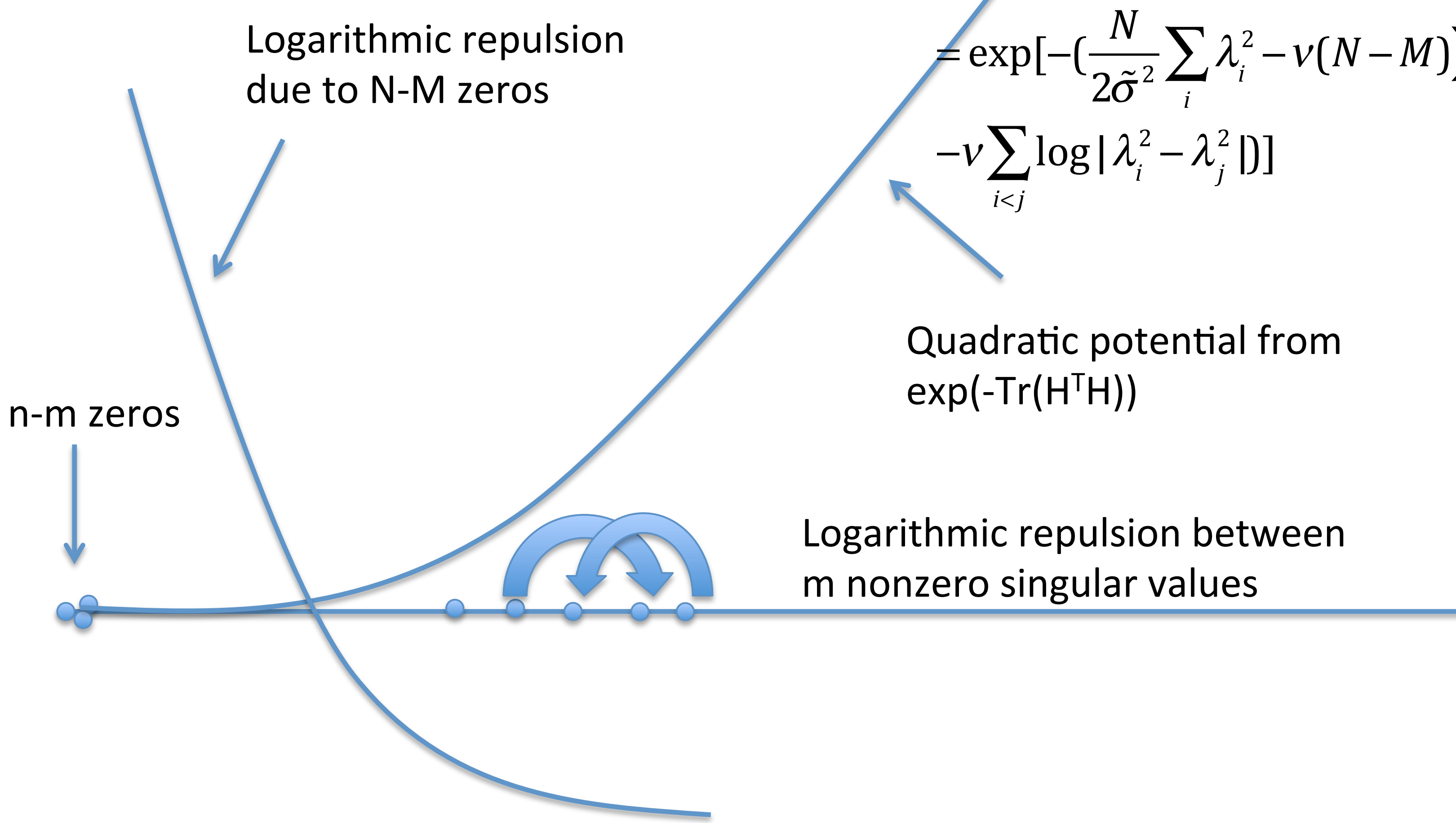
$$H = U\Lambda V^{\dagger} = U\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_M & 0 & \cdots & 0 \end{pmatrix} V^{\dagger}$$

$$P(H)\prod_{ia} dH_{ia} = C\exp[-\frac{N}{2\tilde{\sigma}^2} tr(H^2)]\prod_{ia} dH_{ia}$$

$$\rightarrow Ce^{-\frac{N}{2\tilde{\sigma}^2}\sum_i \lambda_i^2} \prod_{i<j} |\lambda_i^2 - \lambda_j^2|^{\nu} \prod_i \lambda_i^{\nu(N-M+1)-1} (\prod_i d\lambda_i) d\mu(U) d\mu(V)$$

Coulomb of Singular Values

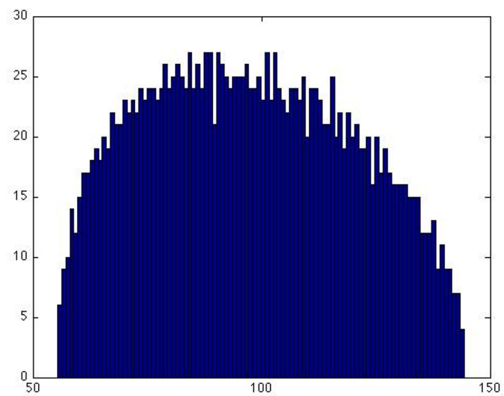
$$
\begin{aligned}
p(\lambda_1,\ldots,\lambda_N) &\propto \exp[-E(\lambda_1,\ldots,\lambda_N)] \\
&= \exp[-(\frac{N}{2\tilde{\sigma}^2}\sum_i \lambda_i^2 - \nu(N-M)\sum_i \log\lambda_i \\
&\quad -\nu\sum_{i<j}\log|\lambda_i^2-\lambda_j^2|)]
\end{aligned}
$$



SVD of `Just Noise'

$$\rho(\lambda)=\frac{a}{\lambda}\sqrt{(\lambda_{\max}^2-\lambda^2)(\lambda^2-\lambda_{\min}^2)}$$
$$where$$
$$\lambda_{\min,\max}=\sigma\sqrt{(M+N)\mp 2\sqrt{MN}}$$

Marcenko
And
Pastur,
1967



Histogram of singular values
For a 2000X10000 random matrix

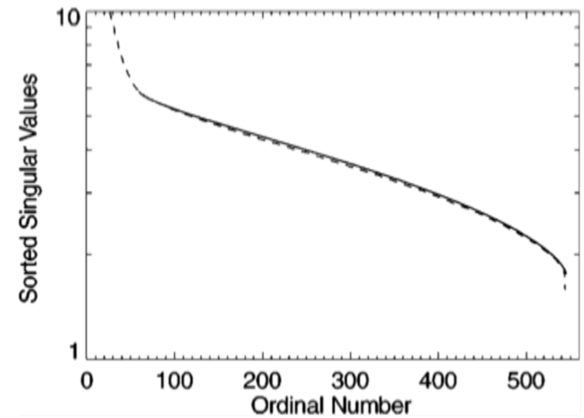
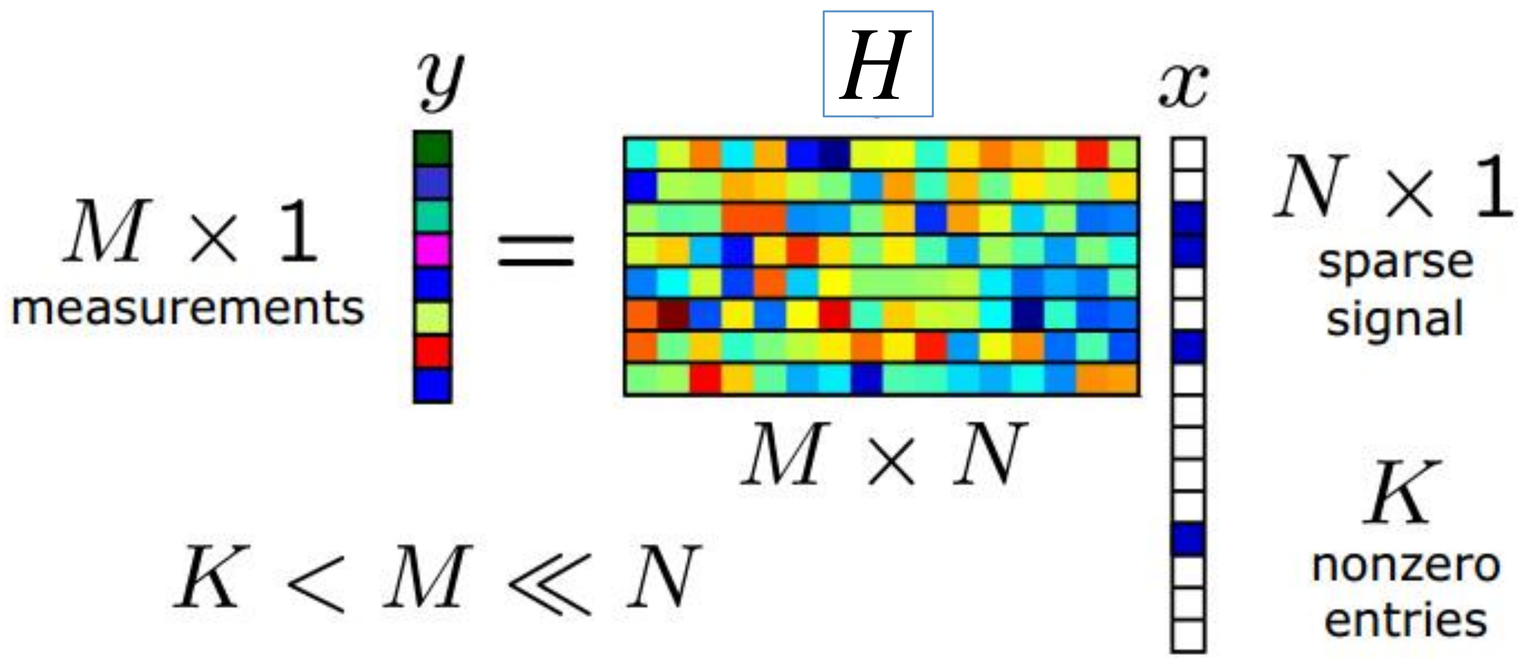


FIG. 3. Comparison of singular values from a SVD of an fMRI data set with the theoretical formula for a noise only matrix.

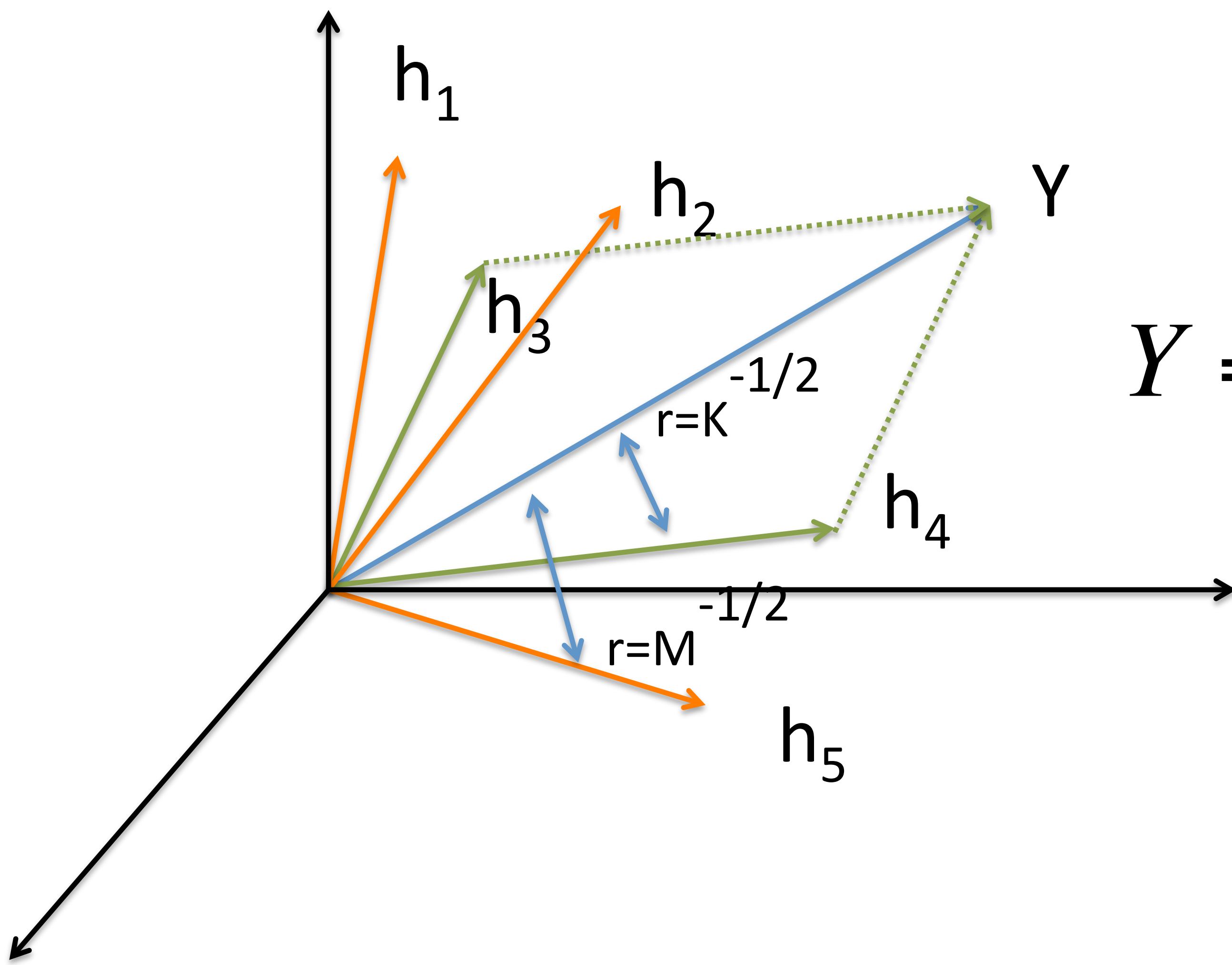
Sengupta and Mitra,
PRE 1999

Gavish, Donoho 2013

Compressed Sensing



Explaining Data



$$Y = \sum_a h_a X_a$$

Convex relaxation

Compressed sensing:

$$y \approx Wx_0, \text{ with } y \in R^M, x_0 \in R^N, M < N,$$

$$\text{but } ||x_0||_0 \le K < M.$$

$$\text{Find } \hat{x} = \arg\min_x ||x||_1 \ \text{ subject to } y = Wx.$$

$$\text{Is } \hat{x} \approx x_0 \text{ when } W \text{ is a random matrix?}$$

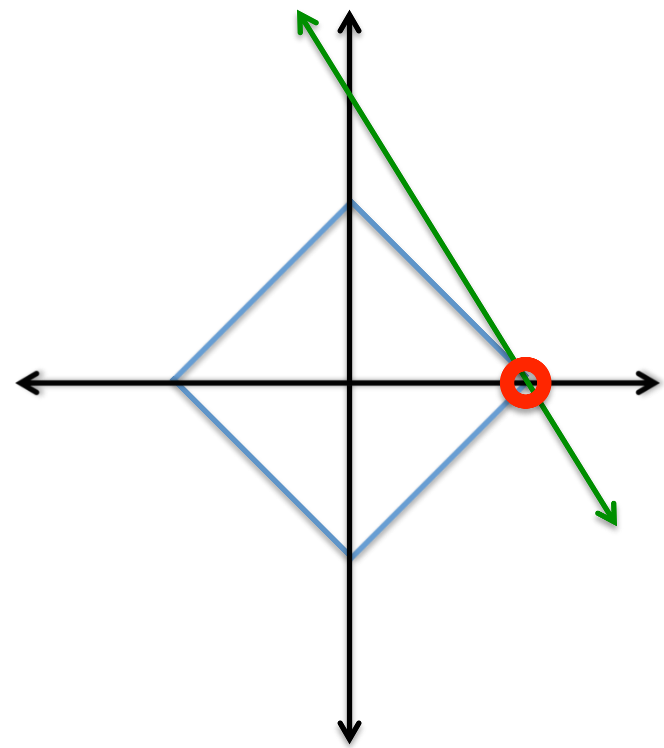
Donoho and Tanner 2009

Sparsity enforced by l_1 norm

L0-L1 relaxation:

Minimize $\| X \|_1$

Subject to $Y = HX$



Works of Candes, Tao
Donoho, from around 2004

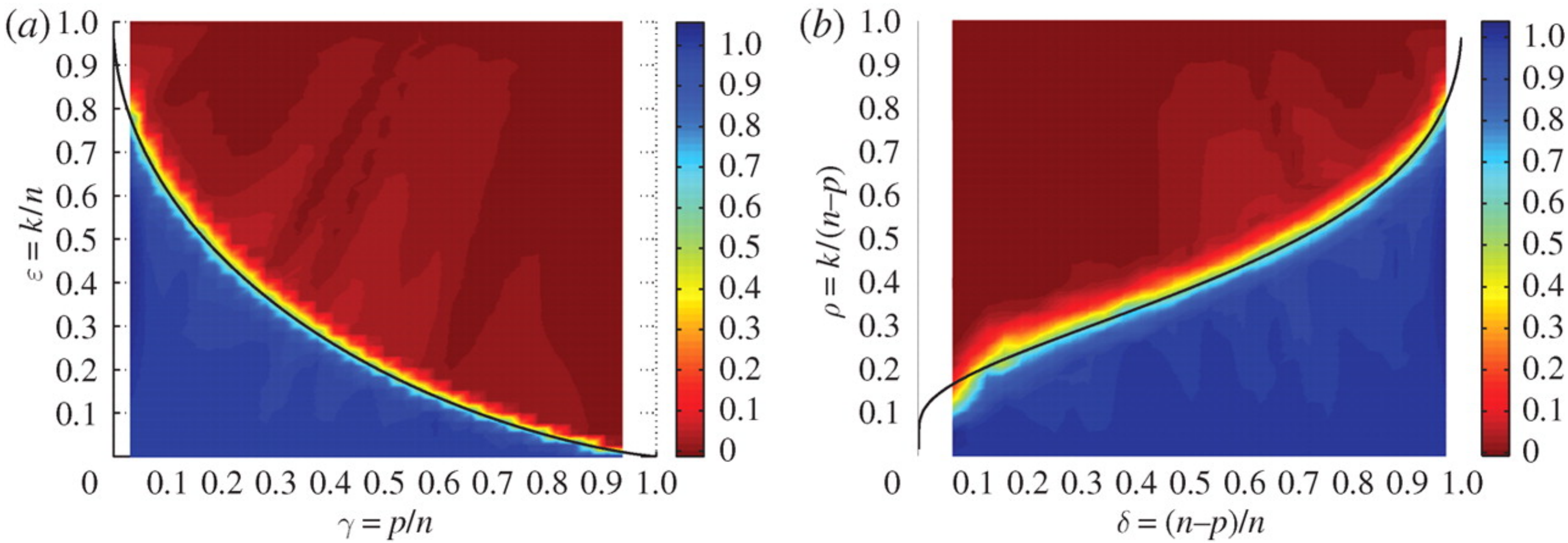
Large Dimensions

The transition happens with

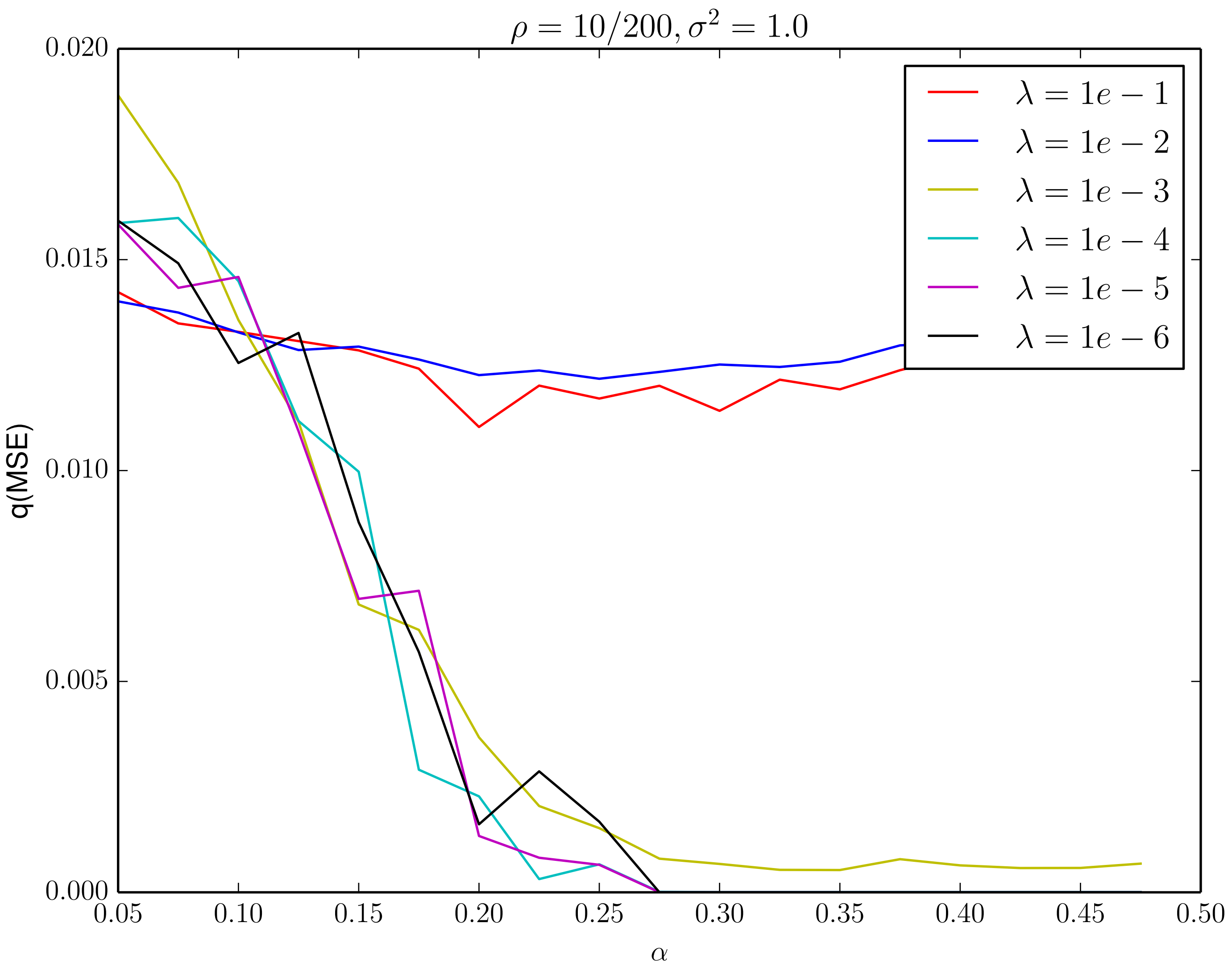
$$N, M, K \to \infty, \text{ holding}$$

$$M / N = \alpha, K / N = \rho \text{ fixed.}$$

Phase Transition in Compressed Sensing by L0-L1 Relaxation



Donoho and Tanner, Phil. Trans. Roy. Soc 2009
Replica Method Analysis: Kabashima
Wadayama and Tanaka, 2009
Ganguly and Sompolinsky, 2010


$\rho = 10/200, \sigma^2 = 1.0$
$\lambda = 1e-1$
$\lambda = 1e-2$
$\lambda = 1e-3$
$\lambda = 1e-4$
$\lambda = 1e-5$
$\lambda = 1e-6$
q(MSE)
0.020
0.015
0.010
0.005
0.000
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
0.50
α

Random-like deep network

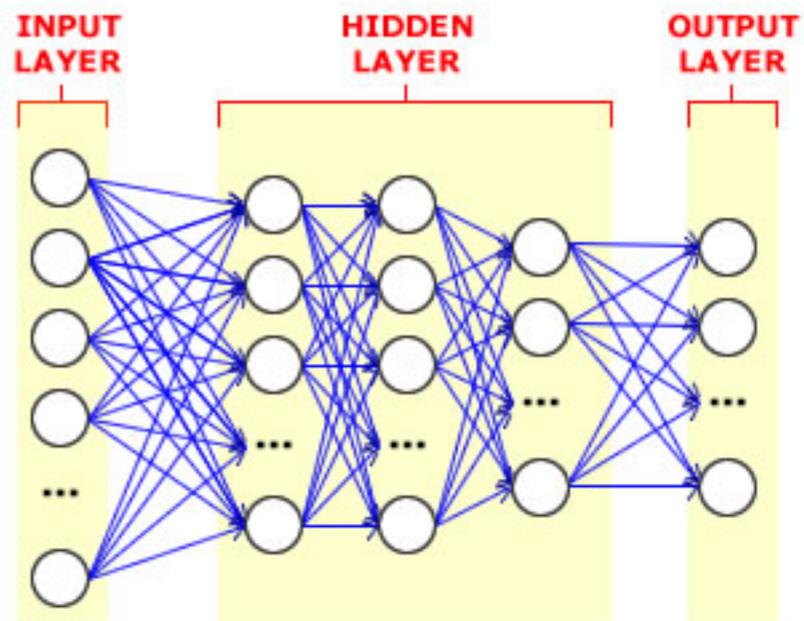
Deep nets

$$h^{(1)}_{M_1\times 1} \approx \sigma(W^{(1)\dagger}_{M_1\times N} x_{N\times 1} + b_1),$$

$$h^{(2)}_{M_2\times 1} \approx \sigma(W^{(2)\dagger}_{M_2\times M_1} h^{(1)}_{M_1\times 1} + b_2)$$

$$\ldots$$

$$h^{(d)}_{M_d\times 1} \approx \sigma(W^{(d)\dagger}_{M_2\times M_1} h^{(d-1)}_{M_1\times 1} + b_d)$$



From http://www.andrewsnoke.com

‘Pre-image’ of the final layer activation

- Fully-connected layers of deep nets

$$h^{(1)} \approx \sigma(W^{(1)\dagger}x \;\; + b_1),$$
$$h^{(2)} \approx \sigma(W^{(2)\dagger}h^{(1)} + b_2)$$
$$\cdots$$
$$h^{(d)} \approx \sigma(W^{(d)\dagger}h^{(d-1)} + b_d)$$

Call this map Φ_f :

$$h^{(d)} = \Phi_f(x)$$

→

What is the preimage of Φ_f ?

Reconstruction :

$$\tilde{x} = \arg\min_x l(\Phi_f(x), h^{(d)}) + R(x)$$

Mahendran and Vedaldi 2015

Alternative ‘generative’ model

Alternative: run the net backwards randomly zeroing some coordinates

Arora, Liang and Ma 2016

$$\tilde{h}^{(d-1)} \approx \tau(W^{(d)}h^{(d)}) \odot n_{d-1}^{drop}$$

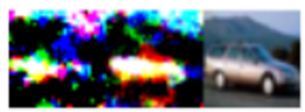
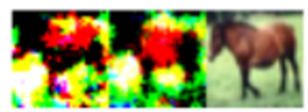
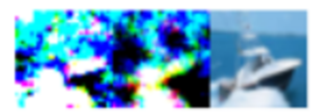
$$\tilde{h}^{(d-2)} \approx \tau(W^{(d-1)}h^{(d-1)}) \odot n_{d-2}^{drop}$$

$$\ldots$$

$$\tilde{x} \approx \tau(W^{(1)}\tilde{h}^{(1)}) \odot n_{0}^{drop},$$

Define a distribution:

$P(\tilde{x} \,|\, h^{(d)})$

Are deep nets reversible when W's are random?

Namely, is $h^{(d)} \approx \Phi_f(\tilde{x})$?

Reversibility in One-layer 'Deep' Net

Run the net backwards

randomly zeroing some coordinates

$\tilde{x} = \gamma \, \mathrm{ReLU}(Wh) \odot n^{drop}$

$\tilde{h} = \mathrm{ReLU}(W^{\dagger}\tilde{x} + b)$

Compute joint distribution $P(\tilde{h}_i, h_i)$

Turns out that $W^{\dagger}\tilde{x} + b$ is nearly Gaussian.

Use that to show $\tilde{h}_i \approx h_i$.

Acknowledgements

Congrats to Spenta and gratitude for all that you have done for me and for Indian Science!

Collaborators: Partha Mitra (CSHL)
Mohammad Ramezanali (Rutgers)
Clement Sire (U. Paul Sabatier)
Mitya Chklovskii (Simons)
Cengiz Pehlevan (Simons)

Funding: NSF, INSPIRE program