

A Measure-Theoretic Dvoretzky Theorem and Applications to Data Science

SEPC in honor of Elizabeth Meckes

Sayan Mukherjee

Duke University
<https://sayanmuk.github.io/>

My last communication

Nov 11, 2020:

Thanks for asking about the workshop, but I don't think I really have anything suitable to talk about. The bigger reason, though, is that I am unfortunately in the middle of a really shitty health crisis, and it seems fairly likely that I'll be pretty much out of commission for a while, so I can't make any commitments at the moment. Sorry. :(

A simple question

What random projections of high-dimensional data or high-dimensional distributions look like ?

1. Exploratory data analysis;

A simple question

What random projections of high-dimensional data or high-dimensional distributions look like ?

1. Exploratory data analysis;
2. Concentration of measure, local theory of Banach spaces;

A simple question

What random projections of high-dimensional data or high-dimensional distributions look like ?

1. Exploratory data analysis;
2. Concentration of measure, local theory of Banach spaces;
3. Proof technique in theory of CS;

A simple question

What random projections of high-dimensional data or high-dimensional distributions look like ?

1. Exploratory data analysis;
2. Concentration of measure, local theory of Banach spaces;
3. Proof technique in theory of CS;
4. Numerical method for scaling algorithms;

A simple question

What random projections of high-dimensional data or high-dimensional distributions look like ?

1. Exploratory data analysis;
2. Concentration of measure, local theory of Banach spaces;
3. Proof technique in theory of CS;
4. Numerical method for scaling algorithms;
5. Proof technique in compressive sensing and theory of deep neural networks;

A simple question

What random projections of high-dimensional data or high-dimensional distributions look like ?

1. Exploratory data analysis;
2. Concentration of measure, local theory of Banach spaces;
3. Proof technique in theory of CS;
4. Numerical method for scaling algorithms;
5. Proof technique in compressive sensing and theory of deep neural networks;
6. Tightest result by Elizabeth – (a) entropy methods and (b) use geometric ideas/arguments.

The problem

1) High dimensional data: X is a random vector in \mathbb{R}^d with $\mathbb{E}X = 0$ and $\mathbb{E}[|X|^2] = \sigma^2 d$

The problem

- 1) High dimensional data: X is a random vector in \mathbb{R}^d with $\mathbb{E}X = 0$ and $\mathbb{E}[|X|^2] = \sigma^2 d$
- 2) Random projection: $\Theta = (\Theta_1, \dots, \Theta_k)$ be a $d \times k$ random projection matrix where

$$X_\theta = (\langle X, \Theta_1 \rangle, \dots, \langle X, \Theta_k \rangle) = \Theta^T X$$

The problem

1) High dimensional data: X is a random vector in \mathbb{R}^d with $\mathbb{E}X = 0$ and $\mathbb{E}[|X|^2] = \sigma^2 d$

2) Random projection: $\Theta = (\Theta_1, \dots, \Theta_k)$ be a $d \times k$ random projection matrix where

$$X_\theta = (\langle X, \Theta_1 \rangle, \dots, \langle X, \Theta_k \rangle) = \Theta^T X$$

3) Marginals are Gaussian: for what values of k is following distance $d(X_\theta, \sigma Z)$ small, where Z is the standard Gaussian random vector in \mathbb{R}^k .

The problem

1) High dimensional data: X is a random vector in \mathbb{R}^d with $\mathbb{E}X = 0$ and $\mathbb{E}[|X|^2] = \sigma^2 d$

2) Random projection: $\Theta = (\Theta_1, \dots, \Theta_k)$ be a $d \times k$ random projection matrix where

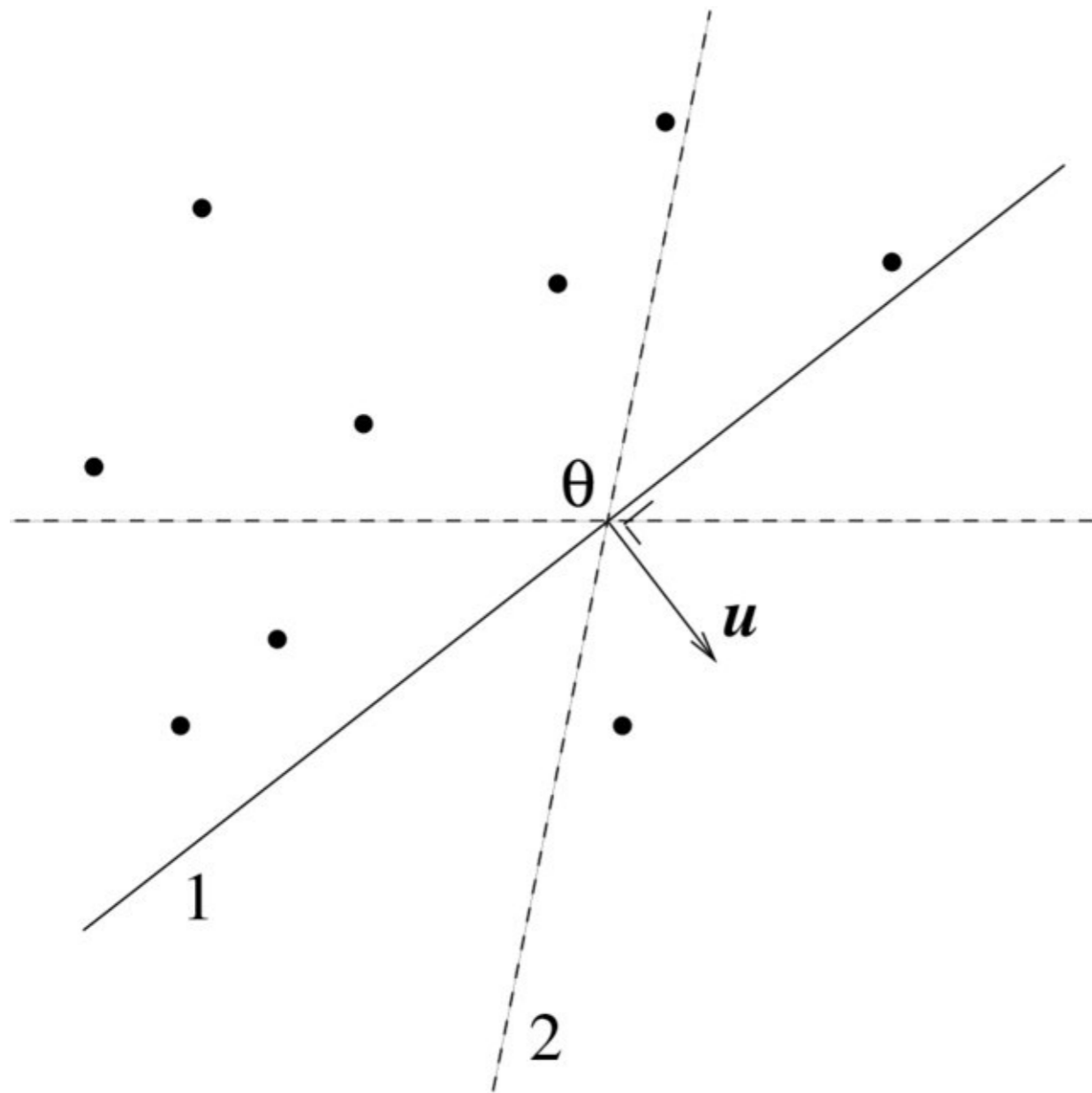
$$X_\theta = (\langle X, \Theta_1 \rangle, \dots, \langle X, \Theta_k \rangle) = \Theta^T X$$

3) Marginals are Gaussian: for what values of k is following distance $d(X_\theta, \sigma Z)$ small, where Z is the standard Gaussian random vector in \mathbb{R}^k .

4) Answer: pre-Elizabeth $k = o(\log(d))$;
Elizabeth $k < \frac{2 \log d}{\log(\log(d))}$.

Data Science to Math

Tukey and data depth



Are there interesting
projection directions ?

Data $D_n = \{x_1, \dots, x_n\}$, and random projection matrix P

$$\text{depth}(\theta, D_n; P) = \frac{1}{n} \#\{i : P^T x_i \geq P^T \theta\}$$

Random projections and Exploratory Data Analysis (EDA)

1) Implementing data depth: Donoho, Huber, Friedman, Kruskal, Stuetzle, Fisherkeller, Diaconis

Random projections and Exploratory Data Analysis (EDA)

- 1) Implementing data depth: Donoho, Huber, Friedman, Kruskal, Stuetzle, Fisherkeller, Diaconis
- 2) Projection pursuit: Donoho, Huber, Friedman, Kruskal, Stuetzle, Fisherkeller, Diaconis

Random projections and Exploratory Data Analysis (EDA)

- 1) Implementing data depth: Donoho, Huber, Friedman, Kruskal, Stuetzle, Fisherkeller, Diaconis
- 2) Projection pursuit: Donoho, Huber, Friedman, Kruskal, Stuetzle, Fisherkeller, Diaconis
- 3) WHP no interesting directions: Diaconis & Freedman, Sudakov

An asymptotic result

Theorem (Diaconis and Freedman)

Let $x_1, \dots, x_n \in \mathbb{R}^d$ and that $n(\nu)$ and $d(\nu)$ goto infinity as $\nu \rightarrow \infty$. There is a $\sigma^2 > 0$ such that for all $\varepsilon > 0$

$$\frac{1}{n} \left| \left\{ j \leq n : \left| |x_j|^2 - \sigma^2 d \right| > \varepsilon d \right\} \right| \xrightarrow{\nu \rightarrow \infty} 0$$

$$\frac{1}{n^2} \left| \left\{ j, k \leq n : \left| \langle x_j, x_k \rangle \right| > \varepsilon d \right\} \right| \xrightarrow{\nu \rightarrow \infty} 0.$$

An asymptotic result

Theorem (Diaconis and Freedman)

Let $x_1, \dots, x_n \in \mathbb{R}^d$ and that $n(\nu)$ and $d(\nu)$ goto infinity as $\nu \rightarrow \infty$. There is a $\sigma^2 > 0$ such that for all $\varepsilon > 0$

$$\frac{1}{n} \left| \left\{ j \leq n : \left| |x_j|^2 - \sigma^2 d \right| > \varepsilon d \right\} \right| \xrightarrow{\nu \rightarrow \infty} 0$$

$$\frac{1}{n^2} \left| \left\{ j, k \leq n : \left| \langle x_j, x_k \rangle \right| > \varepsilon d \right\} \right| \xrightarrow{\nu \rightarrow \infty} 0.$$

Let $\theta \in \mathbb{S}^{d-1}$ be distributed uniformly on the sphere, and $\mu_\nu^\theta = \frac{1}{n} \sum_i \delta_{\langle \theta, x_i \rangle}$. As μ_ν^θ tends to $N(0, \sigma^2)$ weakly in probability.

A quantitative result

Bounded-Lipschitz distance

$$d_{BL}(P, Q) = \sup_f \left| \int dP - \int f dQ \right|, \quad f : \mathbb{R}^k \rightarrow [-1, 1], \text{ one Lipschitz.}$$

A quantitative result

Bounded-Lipschitz distance

$$d_{BL}(P, Q) = \sup_f \left| \int dP - \int f dQ \right|, \quad f : \mathbb{R}^k \rightarrow [-1, 1], \text{ one Lipschitz.}$$

Theorem (Elizabeth)

For projection pursuit most k -dimensional projections of n data points in \mathbb{R}^d are close to Gaussian, when n and d are large and $k = c\sqrt{\log(d)}$.

A quantitative result

Bounded-Lipschitz distance

$$d_{BL}(P, Q) = \sup_f \left| \int dP - \int f dQ \right|, \quad f : \mathbb{R}^k \rightarrow [-1, 1], \text{ one Lipschitz.}$$

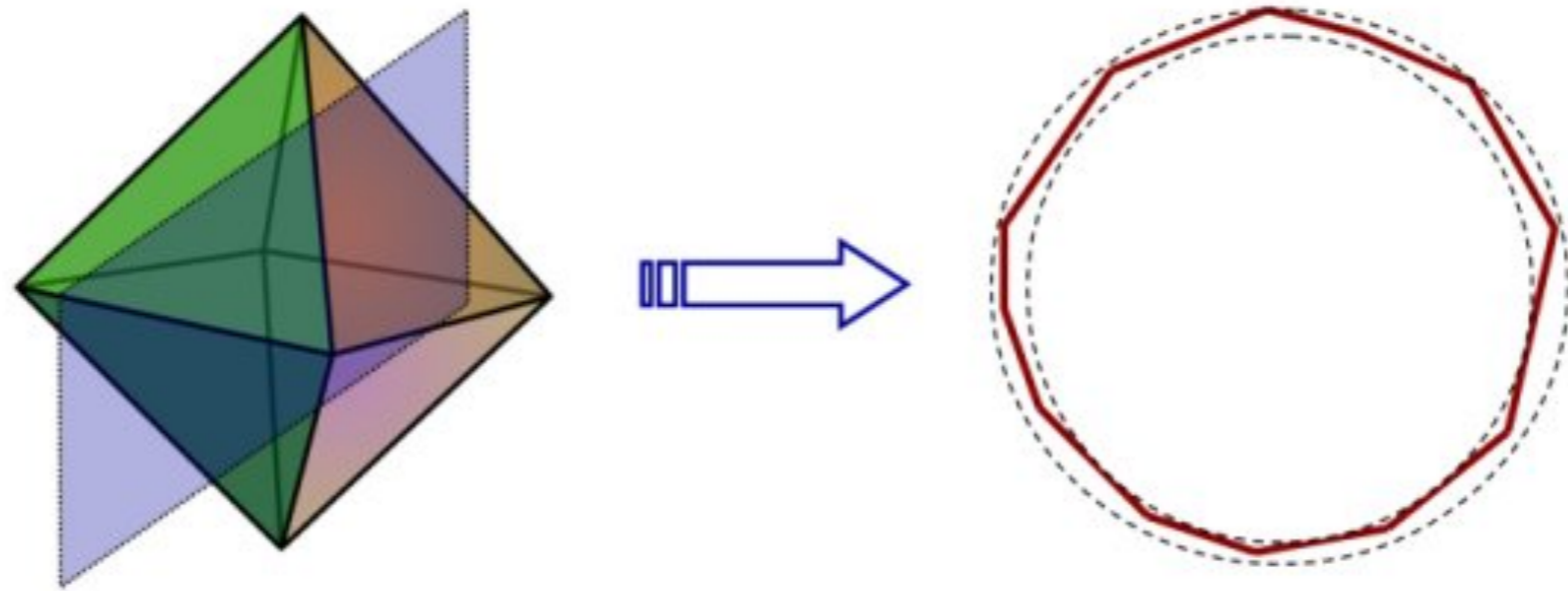
Theorem (Elizabeth)

For projection pursuit most k -dimensional projections of n data points in \mathbb{R}^d are close to Gaussian, when n and d are large and $k = c\sqrt{\log(d)}$.

Stein's method of exchangeable pairs.

Dvoretzky's theorem

A conjecture by Grothendieck: Given a symmetric convex body in Euclidean space of sufficiently high dimensionality, the body will have nearly spherical sections.



Dvoretzky's theorem

Theorem (Dvoretzky)

For every $d \in \mathbb{N}$ and $\varepsilon > 0$ the following holds. Let $|\cdot|$ be the Euclidean norm on \mathbb{R}^d , and let $\|\cdot\|$ be an arbitrary norm. Then there exists a subspace $X \subset \mathbb{R}^d$ with $\dim(X) \geq c(\varepsilon) \log d$, and a number $A > 0$ so that for every $x \in X$

$$A|x| \leq \|x\| \leq (1 + \varepsilon)A|x|.$$

Here, $c(\varepsilon) > 0$ is a constant that depends only on ε .

A central limit theorem for convex sets

Theorem (Klartag)

Let X be a random vector in \mathbb{R}^n with an isotropic log-concave density. There are decreasing sequences ε_n and δ_n for which there exists a subset $\Theta \subset \mathbb{S}^{n-1}$ with $\sigma_{n-1}(\Theta) \geq 1 - \delta_n$ such that for all $\theta \in \Theta$

$$d_{TV}(\langle X, \theta \rangle, Z) < \varepsilon_n, \quad Z \sim N(0, 1).$$

Also

$$\varepsilon_n \leq C \frac{\log \log n + 2}{\log n + 1}, \quad \delta_n \leq \exp(-cn^{.99}).$$

Random subspaces

The Stiefel manifold is the set

$$\mathcal{M}_{d,k} = \left\{ (\theta_1, \dots, \theta_k) : \theta_j \in \mathbb{R}^d, \langle \theta_i, \theta_j \rangle = \delta_{ij} \right\}.$$

$\mathcal{M}_{d,k}$ has a rotation-invariant Haar probability measure.

A measure-theoretic Dvoretzky theorem

Theorem (Elizabeth)

Let X be a random vector in \mathbb{R}^n satisfying

$$\mathbb{E}X = 0, \mathbb{E}|X|^2 = \sigma^2 d, \text{ and } \sup_{\xi \in \mathbb{S}^{d-1}} \mathbb{E}\langle \xi, X \rangle^2 \leq L$$

$$\mathbb{E} \left| |X|^2 \sigma^{-2} - d \right| \leq L \frac{d}{\sqrt{\log(d)}}.$$

A measure-theoretic Dvoretzky theorem

Theorem (Elizabeth)

Let X be a random vector in \mathbb{R}^n satisfying
 $\mathbb{E}X = 0$, $\mathbb{E}|X|^2 = \sigma^2 d$, and $\sup_{\xi \in \mathbb{S}^{d-1}} \mathbb{E}\langle \xi, X \rangle^2 \leq L$
 $\mathbb{E} ||X|^2 \sigma^{-2} - d| \leq L \frac{d}{\sqrt{\log(d)}}.$

For $\theta \in \mathcal{M}_{d,k}$ set X_θ as the projection of X onto the span of θ .
Fix $\delta \in (0, 2)$ and let $k = \delta \frac{\log(d)}{\log(\log(d))}$. Then there is a $c > 0$
depending on δ, L, L' such that for $\varepsilon = \frac{2}{[\log(d)]^c}$, there is a subset
 $\mathcal{I} \subseteq \mathcal{M}_{d,k}$ with $\mathbb{P}[\mathcal{I}^c] \leq C e^{-c' d \varepsilon^2}$, such that for all $\theta \in \mathcal{I}$

$$d_{BL}(X_\theta, \sigma Z) \leq C' \varepsilon.$$

Analogy

- ▶ Given \mathbb{R}^d add structure:
 1. Dvoretzky theorem: the norm
 2. Meckes theorem: the distribution

Analogy

- ▶ Given \mathbb{R}^d add structure:
 1. Dvoretzky theorem: the norm
 2. Meckes theorem: the distribution

- ▶ There is a natural invariant
 1. Dvoretzky theorem: Euclidean norm
 2. Meckes theorem: Gaussian distribution

Math back to Data Science

Compressed sensing

- 1) A signal $X \in \mathbb{R}^d$
- 2) A linear measurement device Θ which is $d \times k$ matrix
- 3) An observation

$$Z = \Theta^T X + \varepsilon, \quad \varepsilon \sim \mathbf{N}(0, \sigma_k^2)$$

Compressed sensing

- 1) A signal $X \in \mathbb{R}^d$
- 2) A linear measurement device Θ which is $d \times k$ matrix
- 3) An observation

$$Z = \Theta^T X + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma_k^2)$$

The question: How many measurements k and what conditions on the noise do we need to recover X ?

Compressed sensing

- 1) A signal $X \in \mathbb{R}^d$
- 2) A linear measurement device Θ which is $d \times k$ matrix
- 3) An observation

$$Z = \Theta^T X + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma_k^2)$$

The question: How many measurements k and what conditions on the noise do we need to recover X ?

1. Linear regression in statistics
2. Multivariate channel in communication systems
3. Signal acquisition in compressed sensing

Algorithms for inference

Not linear algebra ☹

Algorithms for inference

Not linear algebra ☹

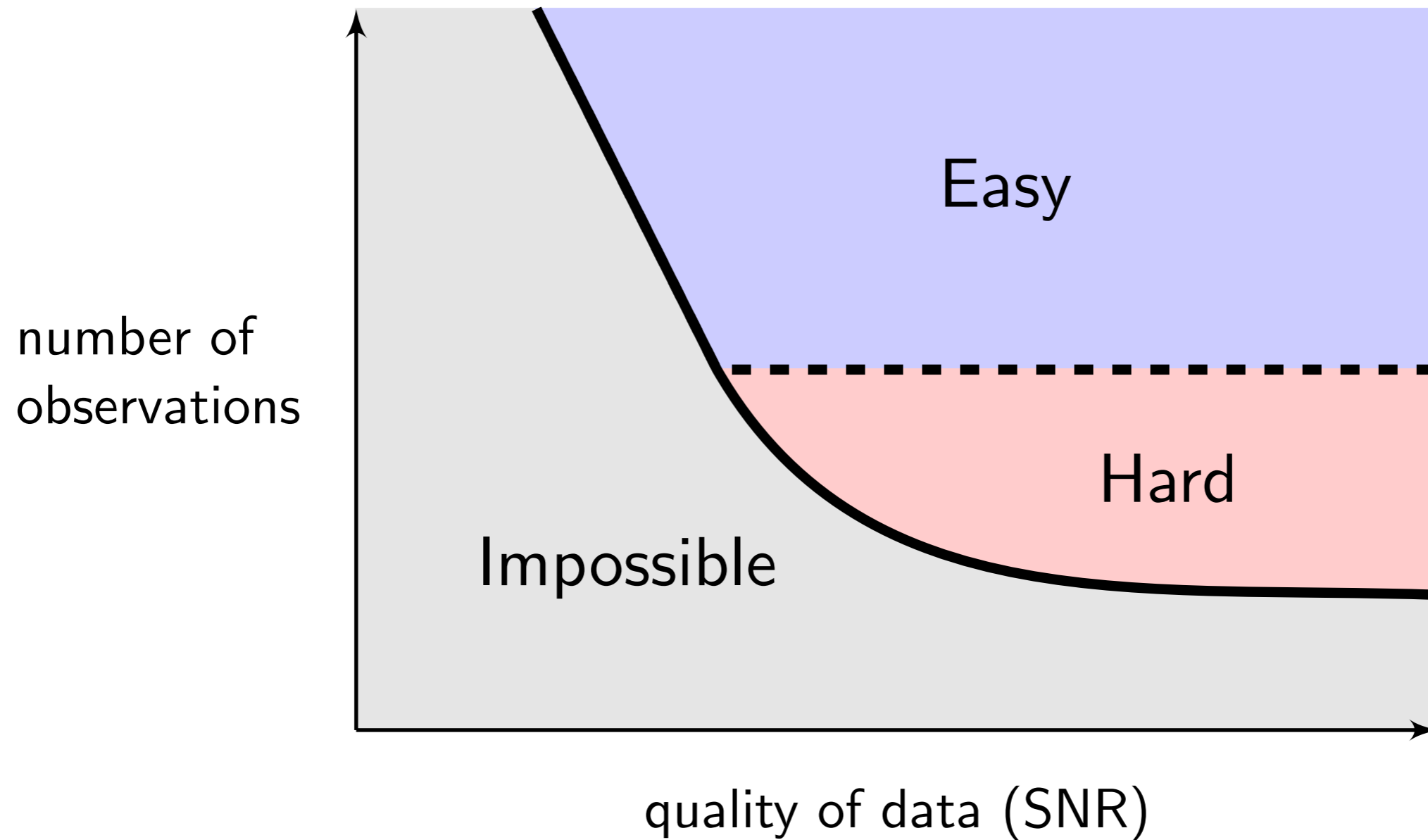
Approximate message-passing (AMP) algorithm

1. Initialize $x^0 = 0$

2. Update

$$\begin{aligned}x^{t+1} &= \eta(\Theta^T \varepsilon^t + x^t), \\ \varepsilon^t &= z - \Theta^T x^t + \frac{1}{\delta} \varepsilon^{t-1} (\eta'(\Theta^T \varepsilon^{t-1} + x^{t-1})).\end{aligned}$$

Typical phase diagram



Algorithms for approximate inference

Variational Inference

- ▶ Expectation consistent framework [Oppor & Winther 2004](#)

Approximate message passing (AMP) with precise analysis via state evolution formalism

- ▶ AMP [Donoho et al. 2009](#), [Bayati & Montanari 2011](#)
- ▶ GAMP [Rangan, 2011](#)
- ▶ S-AMP [Çakmak, Winther, Fleury, 2014](#)
- ▶ O-AMP [Ma & Ping 2016](#)
- ▶ VAMP [Rangan, Schniter, Fletcher, 2016](#)
- ▶ GVAMP [Schniter, Rangan, Fletcher 2016](#)

Long history of related work

Over thirty years of work.

- ▶ Replica method **Parisi, 1980**
- ▶ Free probability theory **Voiculescu, 1990**

Long history of related work

Over thirty years of work.

- ▶ Replica method [Parisi, 1980](#)
- ▶ Free probability theory [Voiculescu, 1990](#)

Wireless communication systems

- ▶ Gaussian case (free probability) [Tse, 1999, Tulino & Verdú, 2004](#)
- ▶ Binary case (replica method) [Kabashima 2003, Tanaka 2004, Tulino & Verdú, 2004](#)

Long history of related work

Over thirty years of work.

- ▶ Replica method [Parisi, 1980](#)
- ▶ Free probability theory [Voiculescu, 1990](#)

Wireless communication systems

- ▶ Gaussian case (free probability) [Tse, 1999, Tulino & Verdú, 2004](#)
- ▶ Binary case (replica method) [Kabashima 2003, Tanaka 2004, Tulino & Verdú, 2004](#)

Replica method and compressed sensing

- ▶ [Guo et al. 2008, Korada & Macris 2010, R. & Gastpar 2012, Wu & Verdu 2012, Krzakala et al. 2012, Donoho et al. 2013, Tulino et al. 2013, Huleihel & Merhav 2016](#)

Very recent progress

Rigorous proofs of replica formulas

- ▶ Linear model, IID Gaussian matrix [R & Pfister. 2016](#)
- ▶ Another proof via spatial coupling [Barbier, Dia, Macris, Krzakala 2016](#)
- ▶ GLM, IID Gaussian matrix [Barbier, Krzakala, Macris, Miolane, Zdeborová. 2017](#)

Very recent progress

Rigorous proofs of replica formulas

- ▶ Linear model, IID Gaussian matrix [R & Pfister. 2016](#)
- ▶ Another proof via spatial coupling [Barbier, Dia, Macris, Krzakala 2016](#)
- ▶ GLM, IID Gaussian matrix [Barbier, Krzakala, Macris, Miolane, Zdeborová. 2017](#)

Focus on multilayer models

- ▶ ML-AMP [Manoel, Krzakala, Mézard, Zdeborová 2017](#)
- ▶ ML-VAMP [Fletcher & Rangan 2017](#)

Compressed sensing and concentration of measure

Reeves and Pfister 2016: The replica prediction is correct for i.i.d. Gaussian measurement matrices provided that the signal distribution, P_X , has bounded fourth moment and satisfies a certain 'single-crossing' property.

Compressed sensing and concentration of measure

Reeves and Pfister 2016: The replica prediction is correct for i.i.d. Gaussian measurement matrices provided that the signal distribution, P_X , has bounded fourth moment and satisfies a certain 'single-crossing' property.

A basic challenge in the proof – control the measure of non-Gaussianness of the conditional distribution of the new measurement.

Compressed sensing and concentration of measure

Reeves and Pfister 2016: The replica prediction is correct for i.i.d. Gaussian measurement matrices provided that the signal distribution, P_X , has bounded fourth moment and satisfies a certain ‘single-crossing’ property.

A basic challenge in the proof – control the measure of non-Gaussianness of the conditional distribution of the new measurement.

Conditional Central Limit Theorems for Gaussian Projections.
Reeves 2016: Adaptation and refinement of results very similar to “measure-theoretic Dvoretzky theorem”.

Compressed sensing and concentration of measure

Reeves and Pfister 2016: The replica prediction is correct for i.i.d. Gaussian measurement matrices provided that the signal distribution, P_X , has bounded fourth moment and satisfies a certain 'single-crossing' property.

A basic challenge in the proof – control the measure of non-Gaussianness of the conditional distribution of the new measurement.

Conditional Central Limit Theorems for Gaussian Projections.
Reeves 2016: Adaptation and refinement of results very similar to “measure-theoretic Dvoretzky theorem”.

Elizabeth – “We long suspected concentration of measure and compressed sensing were somehow linked, these papers make the connection clear.”

Projections and topology

Projections and isometries

Lemma (Johnson-Lindenstrauss)

Fix $0 < \varepsilon < 1$ and $\{x_1, \dots, x_n\} \in \mathbb{R}^d$. If $k \geq \frac{c}{\varepsilon^2} \log n$ then there exists a linear map $\Theta : \mathbb{R}^d \rightarrow \mathbb{R}^k$ such that for all $i \neq j$

$$(1 - \varepsilon)\|x_i - x_j\| \leq \|\Theta(x_i) - \Theta(x_j)\| \leq (1 + \varepsilon)\|x_i - x_j\|.$$

An example: Random projections $\Theta_{ij} \stackrel{iid}{\sim} N(0, 1)$ preserve isometries.

Projections and homology

Given $X = \{x_1, \dots, x_n\} \in \mathbb{R}^d$ construct

$$S(X, \tau) = \bigcup_{i=1}^n B(x_i, \tau), \quad S(X_\theta, \tau') = \bigcup_{i=1}^n B(\Theta^T x_i, \tau'),$$

Projections and homology

Given $X = \{x_1, \dots, x_n\} \in \mathbb{R}^d$ construct

$$S(X, \tau) = \bigcup_{i=1}^n B(x_i, \tau), \quad S(X_\theta, \tau') = \bigcup_{i=1}^n B(\Theta^T x_i, \tau'),$$

with $\Theta \in \mathcal{M}_{d,k}$ for what k and ranges of τ', τ does the following hold (in an interesting way)

$$H_*(S(X, \tau)) \cong H_*(S(X_\Theta, \tau')).$$

Projections and homology

Given $X = \{x_1, \dots, x_n\} \in \mathbb{R}^d$ construct

$$S(X, \tau) = \bigcup_{i=1}^n B(x_i, \tau), \quad S(X_\theta, \tau') = \bigcup_{i=1}^n B(\Theta^T x_i, \tau'),$$

with $\Theta \in \mathcal{M}_{d,k}$ for what k and ranges of τ', τ does the following hold (in an interesting way)

$$H_*(S(X, \tau)) \cong H_*(S(X_\Theta, \tau')).$$

Do random projections:

- 1) Preserve homology ?
- 2) How is this different than preserving isometries ?
- 3) What do random projections do to critical points ?

Projections and genetics

Inference of population structure

A classic problem in biology and genetics is to study population structure.

(1) Does genetic variation in populations follow geography ?

Inference of population structure

A classic problem in biology and genetics is to study population structure.

- (1) Does genetic variation in populations follow geography ?
- (2) Can we infer population histories from genetic variation ?

Inference of population structure

A classic problem in biology and genetics is to study population structure.

- (1) Does genetic variation in populations follow geography ?
- (2) Can we infer population histories from genetic variation ?
- (3) When we associate genetic loci (locations) to disease we need to correct for population structure.

Genetic data

For each individual we have two letters from $\{A, C, T, G\}$ at each polymorphic (SNP) site which is coded as an integer $\{0, 1, 2\}$

$$C_i = \begin{pmatrix} AC \\ \vdots \\ GG \\ \vdots \\ TT \end{pmatrix} \Rightarrow \begin{pmatrix} 1 \\ \vdots \\ 0 \\ \vdots \\ 2 \end{pmatrix} \in \mathbb{R}^{500,000},$$

Genetic data

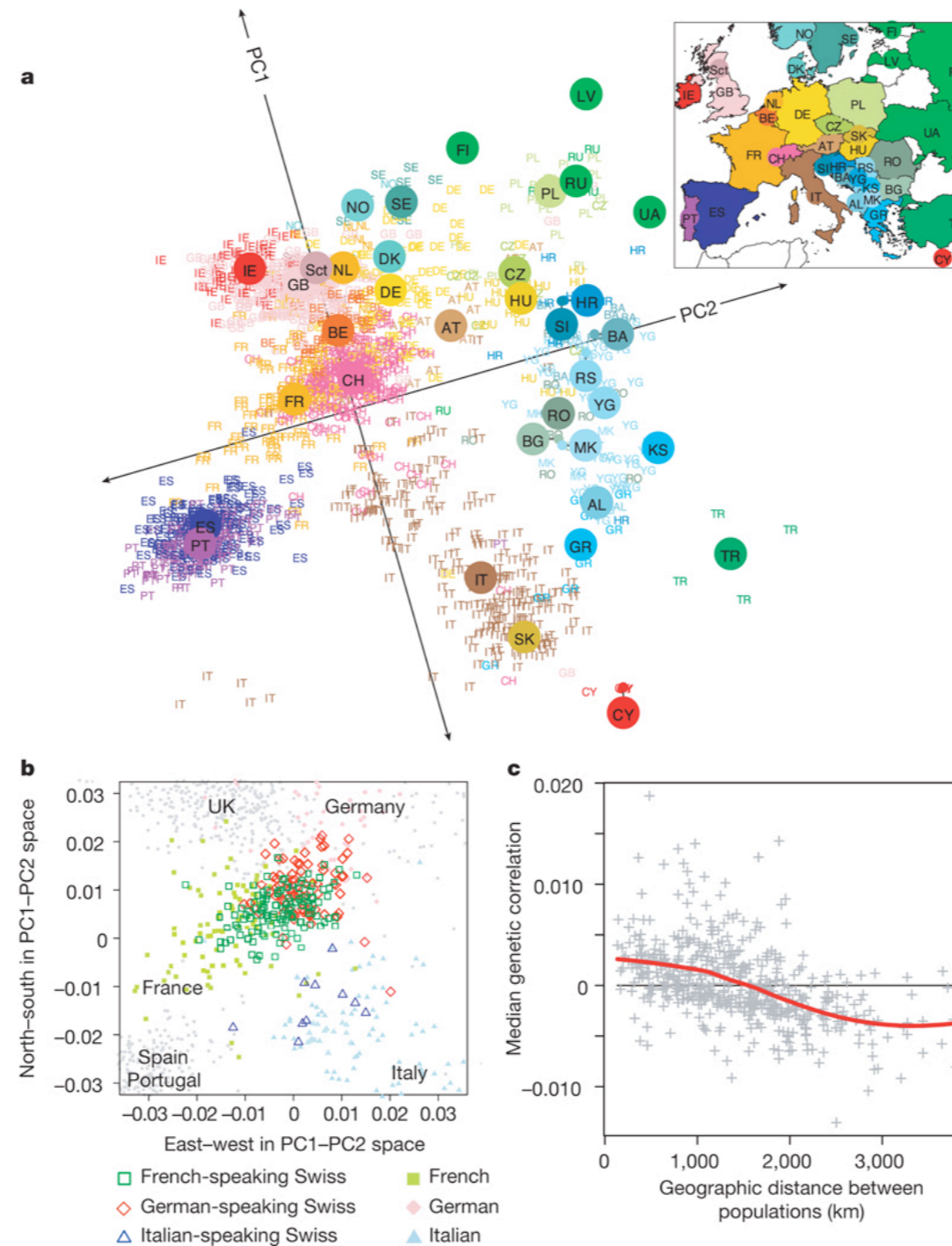
For each individual we have two letters from $\{A, C, T, G\}$ at each polymorphic (SNP) site which is coded as an integer $\{0, 1, 2\}$

$$C_i = \begin{pmatrix} AC \\ \vdots \\ GG \\ \vdots \\ TT \end{pmatrix} \Rightarrow \begin{pmatrix} 1 \\ \vdots \\ 0 \\ \vdots \\ 2 \end{pmatrix} \in \mathbb{R}^{500,000},$$

$$C = [C_1, \dots, C_m].$$

Genetic data encodes population history

From Novembre et al 2008 (Nature)



Popular method

Eigenstrat: Patterson et al 2006 (PLoS Genetics)
Combines principal components analysis and Tracy-Widom theory to infer population structure.

Popular method

Eigenstrat: Patterson et al 2006 (PLoS Genetics)

Combines principal components analysis and Tracy-Widom theory to infer population structure.

$$(1) \quad M_{ij} = \frac{C_{ij} - \hat{\mu}_j}{\sqrt{\frac{\hat{\mu}_j}{2} \left(1 - \frac{\hat{\mu}_j}{2}\right)}} \quad \forall i, j.$$

Popular method

Eigenstrat: Patterson et al 2006 (PLoS Genetics)

Combines principal components analysis and Tracy-Widom theory to infer population structure.

$$(1) \quad M_{ij} = \frac{C_{ij} - \hat{\mu}_j}{\sqrt{\frac{\hat{\mu}_j}{2} \left(1 - \frac{\hat{\mu}_j}{2}\right)}} \quad \forall i, j.$$

$$(2) \quad X = \frac{1}{n} MM'$$

Popular method

Eigenstrat: Patterson et al 2006 (PLoS Genetics)

Combines principal components analysis and Tracy-Widom theory to infer population structure.

$$(1) \quad M_{ij} = \frac{C_{ij} - \hat{\mu}_j}{\sqrt{\frac{\hat{\mu}_j}{2} \left(1 - \frac{\hat{\mu}_j}{2}\right)}} \quad \forall i, j.$$

$$(2) \quad X = \frac{1}{n} MM'$$

(3) Order $\lambda_1, \dots, \lambda_m$ and test for significant eigenvalues using TW statistics

Popular method

Eigenstrat: Patterson et al 2006 (PLoS Genetics)

Combines principal components analysis and Tracy-Widom theory to infer population structure.

$$(1) \quad M_{ij} = \frac{C_{ij} - \hat{\mu}_j}{\sqrt{\frac{\hat{\mu}_j}{2} \left(1 - \frac{\hat{\mu}_j}{2}\right)}} \quad \forall i, j.$$

$$(2) \quad X = \frac{1}{n} MM'$$

(3) Order $\lambda_1, \dots, \lambda_m$ and test for significant eigenvalues using TW statistics

(4) Compute

$$n' = \frac{(m+1) (\sum_i \lambda_i)^2}{((m-1) \sum_i \lambda_i^2) - (\sum_i \lambda_i)^2}.$$

The challenge

Large datasets are being collected (UK Biobank)
 $n \geq 500,000$ and $m \geq 500,000$.

The challenge

Large datasets are being collected (UK Biobank)
 $n \geq 500,000$ and $m \geq 500,000$.

Can we extend Eigenstrat to this data to be run on a standard desktop on the order of minutes?

The challenge

Large datasets are being collected (UK Biobank)
 $n \geq 500,000$ and $m \geq 500,000$.

Can we extend Eigenstrat to this data to be run on a standard desktop on the order of minutes?

Yes: use random projections and the power method:
Fast Principal-Component Analysis Reveals Convergent
Evolution of ADH1B in Europe and East Asia, American
Journal of Human Genetics, 2016.