# Large Deviations for Weakly Interacting Particle Systems

Kavita Ramanan
Brown University

ICTS Colloquium
Bangalore, Dec 3 2018

# Outline of Talk

- PART I: Gibbs measures on $n$-particle configurations
  - Problem Description and Motivation
  - Background on Large Deviations
  - Results
- PART II: Random Projections of High-dimensional measures
  - Random Projections
  - The CLT for Convex Sets
  - Large Deviations for Random Projections of Gibbs Measures

# PART I

Gibbs measures on *n*-particle configurations

# Gibbs measures on $n$-particle configurations

- configuration $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ of $n$ particles in $\mathbb{R}^d$
- subject to
  - a confining potential $V : \mathbb{R}^d \mapsto (-\infty, \infty]$
  - a symmetric pairwise interaction potential $W : \mathbb{R}^d \times \mathbb{R}^d \mapsto (-\infty, \infty]$
- Probability distribution on $n$-particle configurations

$$P_n(d\mathbf{x}_1, ..., d\mathbf{x}_n) \doteq \frac{\exp\left(-\gamma_n H_n(\mathbf{x}_1, ..., \mathbf{x}_n)\right)}{Z_n} \ell(d\mathbf{x}_1) \cdots \ell(d\mathbf{x}_n),$$

where $\ell$ is a non-atomic, $\sigma$-finite measure on $\mathbb{R}^d$
(e.g., Lebesgue measure)
and $Z_n$ is the normalizing constant or partition function

# Gibbs measures on $n$-particle configurations

Recall

$$P_n(d\mathbf{x}_1,...,d\mathbf{x}_n) =: \frac{\exp\left(-\gamma_n H_n(\mathbf{x}_1,...,\mathbf{x}_n)\right)}{Z_n}\ell(d\mathbf{x}_1)\cdots\ell(d\mathbf{x}_n),$$

where $H_n$ is the Hamiltonian defined by

$$H_n(\mathbf{x}^n) \equiv H_n(\mathbf{x}_1,...,\mathbf{x}_n) \quad =: \quad \frac{1}{n}\sum_{i=1}^{n} V(\mathbf{x}_i) + \frac{1}{2n^2}\sum_{i,j=1,i\neq j}^{n} W(\mathbf{x}_i,\mathbf{x}_j)$$

which can be rewritten as

$$H_n(\mathbf{x}^n) = \int_{\mathbb{R}^d} V(\mathbf{x})\, L_n(\mathbf{x}^n; d\mathbf{x}) + \frac{1}{2}\int_{\neq} W(\mathbf{x},\mathbf{y})\, m^n_{\mathbf{x}^n}(d\mathbf{x})\, m^n_{\mathbf{x}^n}(d\mathbf{y}).$$

# Gibbs measures on $n$-particle configurations

Recall

$$P_n\left(d\mathbf{x}_1, ..., d\mathbf{x}_n\right) =: \frac{\exp\left(-\gamma_n H_n\left(\mathbf{x}_1, ..., \mathbf{x}_n\right)\right)}{Z_n} \ell(d\mathbf{x}_1) \cdots \ell(d\mathbf{x}_n),$$

where $H_n$ is the Hamiltonian defined by

$$H_n(\mathbf{x}^n) \equiv H_n\left(\mathbf{x}_1, ..., \mathbf{x}_n\right) \quad =: \quad \frac{1}{n}\sum_{i=1}^{n} V(\mathbf{x}_i) + \frac{1}{2n^2}\sum_{i,j=1, i\neq j}^{n} W(\mathbf{x}_i, \mathbf{x}_j)$$

which can be rewritten as

$$H_n(\mathbf{x}^n) = \int_{\mathbb{R}^d} V\left(\mathbf{x}\right) L_n\left(\mathbf{x}^n; d\mathbf{x}\right) + \frac{1}{2}\int_{\neq} W\left(\mathbf{x}, \mathbf{y}\right) m^n_{\mathbf{x}^n}\left(d\mathbf{x}\right) m^n_{\mathbf{x}^n}\left(d\mathbf{y}\right).$$

where $m^n_{\mathbf{x}^n}\left(\cdot\right)$ is the empirical measure:

$$m^n_{\mathbf{x}^n}\left(\cdot\right) =: \frac{1}{n}\sum_{i=1}^{n} \delta_{\mathbf{x}_i}(\cdot),$$

where $\delta_{\mathbf{y}}$ denotes the Dirac delta mass at $\mathbf{y} \in \mathbb{R}^d$, and
$\int_{\neq}$ represents the integral over $\mathbb{R}^d \times \mathbb{R}^d \setminus \{(x, x) : x \in \mathbb{R}^d\}$

# Motivation: where $P_n$ arises – I

$$P_n(d\mathbf{x}_1, ..., d\mathbf{x}_n) =: \frac{\exp(-\gamma_n H_n(\mathbf{x}_1, ..., \mathbf{x}_n))}{Z_n} \ell(d\mathbf{x}_1) \cdots \ell(d\mathbf{x}_n),$$

- Equilibrium distribution of $n$ interacting Brownian particles at inverse temperature $\gamma_n$
  If $V$, $W$ are sufficiently smooth, then $P_n$ (with $\ell$ as Lebesgue measure) is the invariant distribution of the Markovian (reversible) diffusion on $(\mathbb{R}^d)^n$:

  $$dX_t^n = \nabla H_n(X_t^n)dt + dB_t^n$$

  where $B^n = (B_1, \ldots, B_n)$, with $B_1, \ldots, B_n$ are independent standard $d$-dimensional Brownian motions

# Motivation: where $P_n$ arises – II

Recall

$$P_n(d\mathbf{x}_1, ..., d\mathbf{x}_n) =: \frac{\exp\left(-\gamma_n H_n(\mathbf{x}_1, ..., \mathbf{x}_n)\right)}{Z_n} \ell(d\mathbf{x}_1) \cdots \ell(d\mathbf{x}_n),$$

- When $d = 1$, $V(x) = |x|^2$, and $W(x, y) = \beta \log \frac{1}{|x-y|}$ with $\beta > 0$, then $P_n$ is the law of the $\beta$-ensemble in Random Matrix Theory. Specifically, for
  - $\beta = 1$, $P_n$ is the law of the eigenvalues of the Gaussian Orthogonal Ensemble (GOE) of random symmetric matrices
  - $\beta = 2$, $P_n$ is the law of the eigenvalues of the Gaussian Unitary Ensemble (GUE) of random Hermitian matrices

# Motivation: where $P_n$ arises – II contd.

$$P_n(d\mathbf{x}_1, ..., d\mathbf{x}_n) =: \frac{\exp\left(-\gamma_n H_n(\mathbf{x}_1, ..., \mathbf{x}_n)\right)}{Z_n} \ell(d\mathbf{x}_1) \cdots \ell(d\mathbf{x}_n),$$

- When $d = 2$, and $\mathbb{R}^d \equiv \mathbb{C}$, $\gamma_n = n^2$, $V(x) = |x|^2$ and $W(x, y) = 2 \log \frac{1}{|x-y|}$ then $P_n$ is the law of the (complex) eigenvalues of a random $n \times n$ matrix with iid complex Gaussian entries with covariance $\mathbb{I}_2/2n$, where $\mathbb{I}_2$ is the $2 \times 2$ identity matrix.

# Motivation: where $P_n$ arises – II contd.

$$P_n(d\mathbf{x}_1, ..., d\mathbf{x}_n) =: \frac{\exp\left(-\gamma_n H_n(\mathbf{x}_1, ..., \mathbf{x}_n)\right)}{Z_n} \ell(d\mathbf{x}_1) \cdots \ell(d\mathbf{x}_n),$$

- When $d = 2$, and $\mathbb{R}^d \equiv \mathbb{C}$, $\gamma_n = n^2$, $V(x) = |x|^2$ and $W(x, y) = 2 \log \frac{1}{|x-y|}$ then $P_n$ is the law of the (complex) eigenvalues of a random $n \times n$ matrix with iid complex Gaussian entries with covariance $\mathbb{I}_2/2n$, where $\mathbb{I}_2$ is the $2 \times 2$ identity matrix.
- In the last example if $V$ is non-quadratic, then $P_n$ is the law of the spectrum of random normal matrices (Ameur, Y., Hedenmalm, H. and Makarov, N., 2011)

# Law of the Empirical Measure

• "Mean-field" symmetries of the model suggest the study of $Q_n$, the empirical distribution of the particles

$$m^n_{\mathbf{x}^n}(\cdot) =: \frac{1}{n}\sum_{i=1}^{n} \delta_{\mathbf{x}_i}(\cdot),$$

under the distribution

$$P_n(d\mathbf{x}_1, ..., d\mathbf{x}_n) =: \frac{\exp\left(-\gamma_n H_n(\mathbf{x}_1, ..., \mathbf{x}_n)\right)}{Z_n}\ell(d\mathbf{x}_1)\cdots\ell(d\mathbf{x}_n),$$

Note that under $P_n$, $m^n_{\mathbf{x}^n}$ is a random element taking values in $\mathcal{P}(\mathbb{R}^d)$, the space of probability measures on $\mathbb{R}^d$
So $Q_n$ is a probability measure on $\mathcal{P}(\mathbb{R}^d)$

# Objective: Large Deviations

Goal:

Given a suitable sequence $\gamma_n \to \infty$, establish a large deviation principle (LDP) for $\{Q_n\}$: that is, identify a corresponding suitable sequence $\alpha_n \to \infty$ and function $I : \mathcal{P}(\mathbb{R}^d) \mapsto [0, \infty]$ such that

# Objective: Large Deviations

Goal:

Given a suitable sequence $\gamma_n \to \infty$, establish a large deviation principle (LDP) for $\{Q_n\}$: that is, identify a corresponding suitable sequence $\alpha_n \to \infty$ and function $I : \mathcal{P}(\mathbb{R}^d) \mapsto [0, \infty]$ such that

$$Q_n(A) = P_n(m^n_{x^n} \in A) \approx e^{-\alpha_n I(A)}$$

for all nice subsets $A \subset \mathcal{P}(\mathbb{R}^d)$, where

$$I(A) = \inf_{x \in A} I(x)$$

# Objective: Large Deviations

Goal:

Given a suitable sequence $\gamma_n \to \infty$, establish a large deviation principle (LDP) for $\{Q_n\}$: that is, identify a corresponding suitable sequence $\alpha_n \to \infty$ and function $I : \mathcal{P}(\mathbb{R}^d) \mapsto [0, \infty]$ such that

$$Q_n(A) = P_n(m^n_{x^n} \in A) \approx e^{-\alpha_n I(A)}$$

for all nice subsets $A \subset \mathcal{P}(\mathbb{R}^d)$, where

$$I(A) = \inf_{x \in A} I(x)$$

$\{\alpha_n\}$ is called the speed of the LDP and $I$ is called the rate function

# Large Deviations Principles

- Let $(S, \mathcal{S})$ be a topological space, and $\mathcal{P}(S)$ the space of Borel probability measures on that space.
- Given a sequence of measures $\{\theta_n\} \subset \mathcal{P}(S)$, let $\{\alpha_n\}$ be a sequence of positive real numbers such that $\lim_{n\to\infty} \alpha_n = \infty$, and let $\mathcal{H} : S \to [0, \infty]$ be a rate function (lowersemicontinuous with compact level sets). The sequence $\{\theta_n\}$ is said to satisfy a large deviation principle (LDP) with speed $\{\alpha_n\}$ and rate function $\mathcal{H}$ if for each $E \in \mathcal{B}(S)$,

$$
\begin{aligned}
-\inf_{x\in E^o} \mathcal{H}(x) \leq \liminf_{n\to\infty} \alpha_n^{-1} \log(R_n(E)) &\leq \limsup_{n\to\infty} \alpha_n^{-1} \log(R_n(E)) \\
&\leq -\inf_{x\in \bar{E}} \mathcal{H}(x).
\end{aligned}
$$

# Why Large Deviations Principles?

$$Q_n(A) = P_n(m^n_{x^n} \in A) \approx e^{-\alpha_n I(A)}$$

where

$$I(A) = \inf_{x \in A} I(x)$$

- Large deviations is an asymptotic theory that estimates probabilities of rare events (those with exponentially small probability) and represents them as the solution of a variational problem
- In many cases (when the rate function is strictly convex) the minimizer of the variational representation identifies the limit of the empirical measures
- Large deviations principles often provide information on the behavior of the distribution, conditioned on the rare event

# Why the Topology Matters

Contraction Principle

"One gets LDPs for continuous functionals for free"

If $\{X_n\}$ satisfies an LDP at a speed $\{\alpha_n\}$ on a topological space $(S, \mathcal{S})$ with rate function $I$, and the map $F : (S, \mathcal{S}) \mapsto (S', \mathcal{S}')$ is continuous, then $\{F(X_n)\}$ also satisfies an LDP at a speed $\{\alpha_n\}$ with rate function $J : T \mapsto [0, \infty]$ given by

$$J(y) = \inf\{I(x) : F(x) = y, x \in S\}, \quad y \in T$$

# Recall our Goal

To establish LDPs for the sequence $\{Q_n\}$ of the law of the (random) empirical measures

$$m^n_{\mathbf{x}^n}(\cdot) =: \frac{1}{n}\sum_{i=1}^{n}\delta_{\mathbf{x}_i}(\cdot),$$

under

$$P_n(d\mathbf{x}_1, ..., d\mathbf{x}_n) =: \frac{\exp\left(-\gamma_n H_n(\mathbf{x}_1, ..., \mathbf{x}_n)\right)}{Z_n}\ell(d\mathbf{x}_1)\cdots\ell(d\mathbf{x}_n),$$

where

$$H_n(\mathbf{x}^n) \equiv H_n(\mathbf{x}_1, ..., \mathbf{x}_n) \quad =: \quad \frac{1}{n}\sum_{i=1}^{n} V(\mathbf{x}_i) + \frac{1}{2n^2}\sum_{i,j=1, i\neq j}^{n} W(\mathbf{x}_i, \mathbf{x}_j)$$

# Topologies on the Space $\mathcal{P}(\mathbb{R}^d)$

Recall we are interested in LDPs for the sequence of probability measures $Q_n$ on $\mathcal{P}(\mathbb{R}^d)$

- Let $\mathcal{P}(\mathbb{R}^d)$ be the space of probability measures on $\mathbb{R}^d$, equipped with the weak topology; recall $\mu_n$ converges to $\mu$ in the weak topology if and only if

$$\forall f \in C_b(\mathbb{R}^d), \int_{\mathbb{R}^d} f(\mathbf{x})\mu_n(d\mathbf{x}) \to \int_{\mathbb{R}^d} f(\mathbf{x})\mu(d\mathbf{x}),$$

Let $d_W$ denote the so-called Lévy-Prohorov metric on $\mathcal{P}(\mathbb{R}^d)$ that induces this topology.

# Stronger Topologies on the Space $\mathcal{P}(\mathbb{R}^d)$

Fix a positive, continuous function $\psi : \mathbb{R}^d \to \mathbb{R}^+$ that satisfies the growth condition

$$\lim_{c \to \infty} \inf_{\mathbf{x}:\|\mathbf{x}\|=c} \psi(\mathbf{x}) = \infty.$$

# Stronger Topologies on the Space $\mathcal{P}(\mathbb{R}^d)$

Fix a positive, continuous function $\psi : \mathbb{R}^d \to \mathbb{R}^+$ that satisfies the growth condition

$$\lim_{c\to\infty} \inf_{\mathbf{x}:\|\mathbf{x}\|=c} \psi(\mathbf{x}) = \infty.$$

Define

$$\mathcal{P}_\psi(\mathbb{R}^d) \doteq \left\{ \mu \in \mathcal{P}(\mathbb{R}^d) : \int_{\mathbb{R}^d} \psi(\mathbf{x})\, \mu(d\mathbf{x}) < +\infty \right\}.$$

and endow it with the metric

$$d_\psi(\mu,\nu) \doteq d_w(\mu,\nu) + \left| \int_{\mathbb{R}^d} \psi(\mathbf{x})\mu(d\mathbf{x}) - \int_{\mathbb{R}^d} \psi(\mathbf{x})\nu(d\mathbf{x}) \right|,$$

where $d_w$ is the Lévy-Prohorov metric that topologizes weak convergence.

# Stronger Topologies on the Space $\mathcal{P}(\mathbb{R}^d)$

Fix a positive, continuous function $\psi : \mathbb{R}^d \to \mathbb{R}^+$ that satisfies the growth condition

$$\lim_{c\to\infty} \inf_{\mathbf{x}:||\mathbf{x}||=c} \psi(\mathbf{x}) = \infty.$$

Define

$$\mathcal{P}_\psi(\mathbb{R}^d) \doteq \left\{ \mu \in \mathcal{P}(\mathbb{R}^d) : \int_{\mathbb{R}^d} \psi(\mathbf{x})\, \mu(d\mathbf{x}) < +\infty \right\}.$$

and endow it with the metric

$$d_\psi(\mu, \nu) \doteq d_w(\mu, \nu) + \left| \int_{\mathbb{R}^d} \psi(\mathbf{x})\mu(d\mathbf{x}) - \int_{\mathbb{R}^d} \psi(\mathbf{x})\nu(d\mathbf{x}) \right|,$$

where $d_w$ is the Lévy-Prohorov metric that topologizes weak convergence.

Note: The choice of $\psi(\mathbf{x}) = ||\mathbf{x}||^p$ for some $p \in [1, \infty)$ coincides with the $p$-Wasserstein topology

# Results for Case 1: $\gamma_n/n \to \infty$

## Dupuis, Laschos and R '15

Under suitable boundedness and growth conditions on $V$ and $W$, and an additional approximability assumption, the sequence $\{Q_n\}$ satisfies an LDP on $\mathcal{P}(\mathbb{R}^d)$ (respectively $\mathcal{P}^\psi(\mathbb{R}^d)$) with rate function $\mathcal{J}_*$ (respectively, $\mathcal{J}_*^\psi$), where

$$\mathcal{J}_* := \mathcal{J}(\mu) - \inf_{\mu \in \mathcal{P}(\mathbb{R}^d)} \mathcal{J}(\mu),$$

and

$$\mathcal{J}_*^\psi := \mathcal{J}(\mu) - \inf_{\mu \in \mathcal{P}^\psi(\mathbb{R}^d)} \mathcal{J}(\mu),$$

with $\mathcal{J} : \mathcal{P}(\mathbb{R}^d) \to (-\infty, \infty]$, given by

$$\mathcal{J}(\mu) \doteq \frac{1}{2} \int_{\mathbb{R}^d \times s\mathbb{R}^d} (V(\mathbf{x}) + V(\mathbf{y}) + W(\mathbf{x}, \mathbf{y}))\, \mu(d\mathbf{x})\, \mu(d\mathbf{y})$$

# Results for Case 2: $\gamma_n = n$

## Classical Special Case: $W = 0$

$$P_n(d\mathbf{x}_1, \ldots, d\mathbf{x}_n) =: \frac{\exp\left(-\gamma_n H_n(\mathbf{x}_1, \ldots, \mathbf{x}_n)\right)}{Z_n} \ell(d\mathbf{x}_1) \cdots \ell(d\mathbf{x}_n),$$

Particles are iid with marginal distribution

$$\theta(d\mathbf{x}) \propto \exp(-\gamma_n V(\mathbf{x}))$$

### Sanov's Theorem

Then $\{Q_n\}$ satisfies an LDP with rate function $\mathcal{R}$, where $\mathcal{R}(\cdot|\theta)$ is the relative entropy functional: for $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$,

$$\mathcal{R}(\mu|\nu) \doteq \begin{cases} \int_{\mathbb{R}^d} \frac{d\mu}{d\nu}(\mathbf{x}) \log\left(\frac{d\mu}{d\nu}(\mathbf{x})\right) \nu(d\mathbf{x}), & \text{if } \mu \ll \nu, \\ \infty, & \text{otherwise,} \end{cases}$$

where $\mu \ll \nu$ denotes that $\mu$ is absolutely continuous with respect to $\nu$.

# Results II: $\gamma_n = n$

## Theorem (Dupuis, Laschos and R '15)

*Under suitable boundedness and growth conditions on $V$ and $W$, and an additional approximability assumption, the sequence $\{Q_n\}$ satisfies an LDP with rate function $\mathcal{I}_*$ (respy, $\mathcal{I}_*^\psi$ on $\mathcal{P}(\mathbb{R}^d)$ (respy, $\mathcal{P}^\psi(\mathbb{R}^d)$), where*

$$\mathcal{I}_*(\mu) = \mathcal{I}(\mu) - \inf_{\mu \in \mathcal{P}(\mathbb{R}^d)} \mathcal{I}(\mu),$$

*and*

$$\mathcal{I}_*^\psi(\mu) = \mathcal{I}(\mu) - \inf_{\mu \in \mathcal{P}^\psi(\mathbb{R}^d)} \mathcal{I}(\mu),$$

*with*

$$\mathcal{I}(\mu) \doteq \mathcal{R}\left(\mu | e^{-V} \ell\right) + \mathcal{W}(\mu),$$

*where $\mathcal{R}(\cdot|\cdot)$ is the relative entropy functional and*

$$\mathcal{W}(\mu) = \frac{1}{2} \int_{\mathbb{R}^d \times \mathbb{R}^d} W(\mathbf{x}, \mathbf{y}) \mu(d\mathbf{x}) \mu d\mathbf{y}).$$

# Prior Work

- Specific results for random matrices were obtained by Ben-Arous-Guionnet (1997), Ben-Arous-Zeitouni (1998), Petz and Hievi (2998) and A. Hardy (2012)
- In this generality, LDPs with respect to the weak topology were obtained when $\ell$ is Lebesgue measure and $\gamma_n/n \log n \to \infty$ in work by Chafaï, Gozlan and Zitt (2014).
- Our work significantly extends this work by allowing for
    - (i) more general measures $\ell$
      (ii) arbitrary $\gamma_n/n \to \infty$
      (iii) considering stronger topologies
- Our methodology is also different (based on weak convergence methods for large deviations) and based on a common framework

# PART II

Random Projections of High-Dimensional Measures

# Motivation and Context

Theme:

Study high-dimensonal objects
by looking at their
(random) lower-dimensional projections

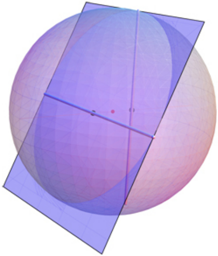
# First Motivation

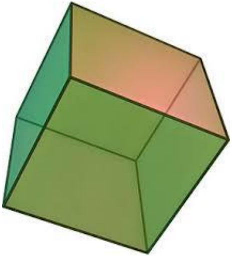
- Asymptotic (Convex) Geometry : study of the geometry of convex bodies in high dmensions
- The presence of high dimensions forces certain regularity on the geometry of convex bodies that has a probabilistic flavor

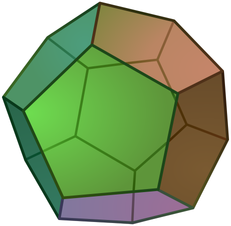
# High Dimension Imposes Regularity

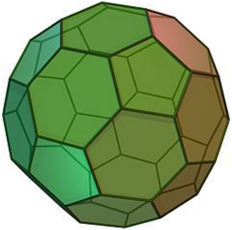

Picture Credit: R. Vershynin

# Asymptotic Convex Geometry









# Motivation and Context:

Theme:
Study high-dimensonal objects
by looking at their
(random) lower-dimensional projections

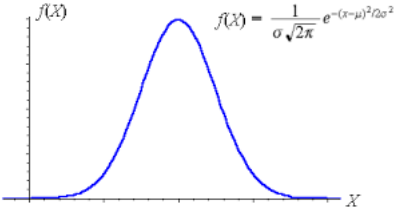
Second Motivation
High-dimensional Probability and Statistics
low-dimensional projections are of relevance in
e.g., sparse recovery, information retrieval, statistics
**(although will not discuss these applications)**

# A Striking Result

Conjectured by
Anttila, Ball and Perissanski (2003); Brehm and Voigt (2000);
(Diaconis & Freedman, Klartag, Meckes-Meckes, Bobkov, ...)

## The CLT for Convex Sets

Let $X^{(n)}$ be uniformly distributed on a convex body in $\mathbb{R}^n$
Then $\langle X^{(n)}, \theta^{(n)} \rangle_n$ is approximately Gaussian for almost all $\theta^{(n)}$ on $S^{n-1}$



# A Precise Statement of one such result

- A probability measure $\mu^n$ (or its density) on $\mathbb{R}^n$ is said to be isotropic if it satisfies

$$\int_{\mathbb{R}^n} x\mu(dx) = 0 \text{ and } \int_{\mathbb{R}^n} x_i x_j \mu(dx) = \delta_{ij}.$$

- A convex body $K$ is said to be isotropic if the normalized uniform measure on the body is isotropic

**Theorem (CLT for Convex Sets; Klartag '07)**

*There exist sequences $\varepsilon_n \downarrow 0$, $\delta_n \downarrow 0$ for which the following holds: if $K \subset \mathbb{R}^n$ is an isotropic convex body and $X^{(n)}$ is the random vector that is distributed uniformly in $K$, then there exists a subset $\mathbb{S} \subset \mathcal{S}^{n-1}$ with $\sigma_{n-1}(\mathbb{S}) \geq 1 - \delta_n$ such that*

$$d_{TV}\left(\langle X^{(n)}, \theta^{(n)} \rangle_n, Z\right) \leq \varepsilon_n$$

*for all $\theta^{(n)} \in \mathbb{S}$, where $Z \sim \mathcal{N}(0,1)$ and $\sigma_{n-1}$ is the inv. measure on $\mathcal{S}^{n-1}$*

# Several Extensions of the CLT

- CLT can be extended to high-dimensional measures that satisfy a certain concentration condition (Meckes-Meckes)
- And it can also be extended to multidimensional projections (Klartag, Meckes, etc.)

# An Implication of Such a Result

Projection-Pursuit Algorithm
Kruskal (1969)
Friedman and Tukey (1974)
Diaconis and Friedman (1987)

1. Finding the “most interesting” possible projections in high-dimensional data
2. “Most interesting” are those that deviate more from a normal distribution

# A Natural Question

- Fluctuation analysis suggests that
  "Typical projections of certain high-dimensional measures that satisfy a geometric condition behave like sums of i.i.d. random variables"
- We know from a classical result called **Cramér's theorem** that sums of i.i.d. random variables satisfy large deviations principles (LDP).

# A Natural Question

- Fluctuation analysis suggests that
  "Typical projections of certain high-dimensional measures that satisfy a geometric condition behave like sums of i.i.d. random variables"
- We know from a classical result called **Cramér's theorem** that sums of i.i.d. random variables satisfy large deviations principles (LDP).

**Question**

Can we establish LDPs for random projections of high-dimensional measures?

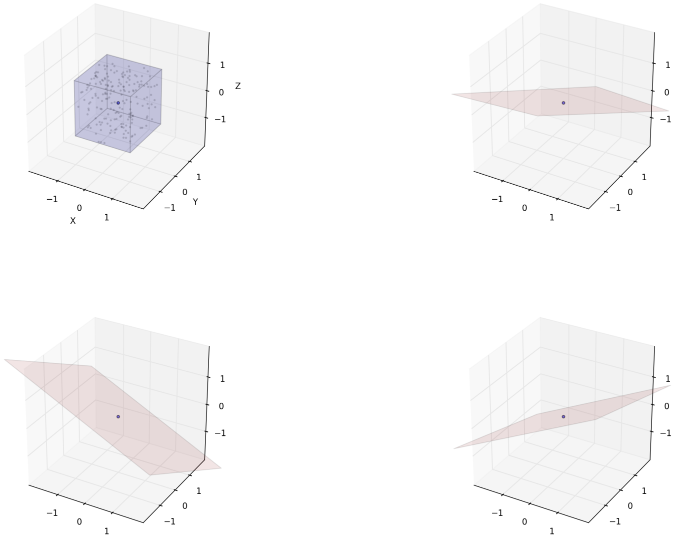
# A Natural Question

- Fluctuation analysis suggests that
  "Typical projections of certain high-dimensional measures that satisfy a geometric condition behave like sums of i.i.d. random variables"
- We know from a classical result called **Cramér's theorem** that sums of i.i.d. random variables satisfy large deviations principles (LDP).

## Question

Can we establish LDPs for random projections of high-dimensional measures?
May more easily distinguish the high-dimensional objects
Would lead to an LDP of a geometric flavour

# Random Orthonormal Bases



# The Stiefel Manifold

The Stiefel manifold of orthonormal $k$-frames in $\mathbb{R}^n$, $k < n$,

$$\mathbb{V}_{n,k} \doteq \{A \in \mathbb{R}^{n \times k} : A^T A = I_k\},$$

where $I_k$ is the $k \times k$ identity matrix.

Random orthonormal frames/bases are chosen with respect to the invariant measure on the (compact) Stiefel manifold.

# Random Projections onto Random subspaces

- Suppose $X^{(n)}$ is a random vector in $\mathbb{R}^n$ and

$$\mathbf{A}_{n,k} = [\mathbf{A}_{n,k}(i,j)]_{i=1,\ldots,n;j=1,\ldots,k}$$

  is an $n \times k$ matrix drawn from the Haar measure on the Stiefel manifold $\mathbb{V}_{n,k}$, independent of $X^{(n)}$.
- Then the coordinates of the projection of $X^{(n)}$ onto the $k$-dimensional subspace with basis determined by $\mathbf{A}_{n,k}$ is given by

$$W_A^{(n)} \doteq A_{n,k}^T X^{(n)}.$$

# Results on Multidimensional Projections

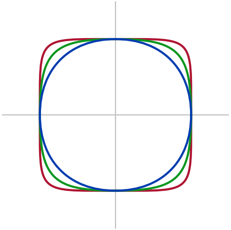
**Assumption: The ATS (Asymptotic Thin-Shell) Condition**

Suppose $||X^{(n)}||_2/\sqrt{n}$ satisfies an LDP in $\mathbb{R}$ with speed $n$ and rate function $J_X : \mathbb{R} \mapsto [0, \infty)$.

**Theorem (Kim and R '17)**

*Suppose the sequence $\{X^{(n)}\}$ satisfies the ATS condition. Then $W_A^{(n)}, n \in \mathbb{N}$, where recall $W_A^{(n)} \doteq A_{n,k}^T X^{(n)}$, satisfies an LDP at speed $n$ with a rate function that can be written explicitly in terms of $J_x$.*

# A natural sequence – $\ell^p$ Balls in $\mathbb{R}^n$



$$\begin{aligned} \|x\|_{n,p} &\doteq \left(x_1^p + \cdots + x_n^p\right)^{1/p} \\ \mathbb{B}_{n,p} &\doteq \{x \in \mathbb{R}^n : \|x\|_{n,p} \leq 1\}. \end{aligned}$$

Let $\mathbf{X}^{(\mathbf{n},\mathbf{p})}$ be uniformly distributed on the $\ell^p$ ball $\mathbb{B}_{\mathbf{n},\mathbf{p}}$.
Then the renormalized scalar projection is given by

$$\mathfrak{W}_\Theta^{(n,p)} \doteq \frac{n^{\mathbf{1}/\mathbf{p}}}{n^{1/2}} \langle \mathbf{X}^{(\mathbf{n},\mathbf{p})}, \Theta^{(n)} \rangle_n.$$

Scaling rationale: "typical" coordinate of $\mathbf{X}^{(\mathbf{n},\mathbf{p})}$ is $\sim n^{-\mathbf{1}/\mathbf{p}}$ and "typical" coordinate of $\Theta^{(n)}$ is $\sim n^{-1/2}$.

# Results: Examples that verify the ATS Condition

$\ell_p$-balls (Kim and R '17)

I. $X^{(n)}$ uniformly distributed on a suitably normalized $\ell_p$ ball in $\mathbb{R}^n$ for $p \geq 2$

# Results: Examples that verify Assumption ATS

The Gibbs Case (Kim and R '17)
II. $X^{(n)} \sim P_n$ (with $d = 1$, $\gamma_n = n$)

# Results: Examples that verify Assumption ATS

The Gibbs Case (Kim and R '17)
II. $X^{(n)} \sim P_n$ (with $d = 1$, $\gamma_n = n$)

Comments on the Proof

- Recall that the ATS condition entails showing that $\{||X^{(n)}||_2/\sqrt{n}\}$ satisfies an LDP
- Note $||x^{(n)}||_2/\sqrt{n}$ can be written in terms of the empirical measure:

$$\frac{||\mathbf{x}^{(n)}||_2}{\sqrt{n}} = \left( \int y^2 m^n_{\mathbf{x}^n}(dy) \right)^{1/2}.$$

# Results: Examples that verify Assumption ATS

The Gibbs Case (Kim and R '17)
II. $X^{(n)} \sim P_n$ (with $d = 1$, $\gamma_n = n$)

Comments on the Proof

- Recall that the ATS condition entails showing that $\{||X^{(n)}||_2/\sqrt{n}\}$ satisfies an LDP
- Note $||x^{(n)}||_2/\sqrt{n}$ can be written in terms of the empirical measure:

$$\frac{||\mathbf{x}^{(n)}||_2}{\sqrt{n}} = \left(\int y^2 m^n_{\mathbf{x}^n}(dy)\right)^{1/2}.$$

  The map $m^n_{\mathbf{x}^n} \mapsto \frac{||\mathbf{x}^n||_2}{\sqrt{n}}$ is continuous in the 2-Wasserstein topology.
- The ATS condition then follows from the LDP for the law $Q_n$ of $m^n_{\mathbf{X}^{(n)}}$ (with $\gamma_n = n$) in the 2-Wasserstein topology and the contraction principle.

# Results: Examples that verify Assumption ATS

The Gibbs Case (Kim and R '17)
II. $X^{(n)} \sim P_n$ (with $d = 1$, $\gamma_n = n$)

Comments on the Proof

- Recall that the ATS condition entails showing that $\{||X^{(n)}||_2/\sqrt{n}\}$ satisfies an LDP
- Note $||x^{(n)}||_2/\sqrt{n}$ can be written in terms of the empirical measure:

$$\frac{||\mathbf{x}^{(n)}||_2}{\sqrt{n}} = \left(\int y^2 m^n_{\mathbf{x}^n}(dy)\right)^{1/2}.$$

  The map $m^n_{\mathbf{x}^n} \mapsto \frac{||\mathbf{x}^n||_2}{\sqrt{n}}$ is continuous in the 2-Wasserstein topology.
- The ATS condition then follows from the LDP for the law $Q_n$ of $m^n_{\mathbf{X}^{(n)}}$ (with $\gamma_n = n$) in the 2-Wasserstein topology and the contraction principle.
  This justifies the need for establishing the LDP in stronger topologies !!

# Summary

- Laws of interacting particles (equivalently, non-product measures in high-dimensions) arise in many applications, ranging from statistics to statistical mechanics
- The study of their large deviations behavior is of general interest and leads to the study of certain associated variational problems
- The rate function in large deviations provides interesting qualitative information about the system