A CLT for Wishart Tensors

Dan Mikulincer

Weizmann Institute of Science

Wishart Tensors

Let $\{X_i\}_{i=1}^d$ be *i.i.d.* copies of an isotropic random vector $X \sim \mu$ in $\mathbb{R}^n$. Denote by $\mathcal{W}_{n,d}^p(\mu)$ the law of

$$\frac{1}{\sqrt{d}} \sum_{i=1}^{d} \left(X_i^{\otimes p} - \mathbb{E}\left[X_i^{\otimes p} \right] \right).$$

We are interested in the behavior as $d \to \infty$. Specifically, when is it true that $\mathcal{W}_{n,d}^p(\mu)$ is approximately Gaussian?

Wishart Tensors

Let $\{X_i\}_{i=1}^d$ be *i.i.d.* copies of an isotropic random vector $X \sim \mu$ in $\mathbb{R}^n$. Denote by $\mathcal{W}_{n,d}^p(\mu)$ the law of

$$\frac{1}{\sqrt{d}} \sum_{i=1}^{d} \left(X_i^{\otimes p} - \mathbb{E}\left[X_i^{\otimes p} \right] \right).$$

We are interested in the behavior as $d \to \infty$. Specifically, when is it true that $\mathcal{W}_{n,d}^p(\mu)$ is approximately Gaussian?

Technicalities

$\mathcal{W}^p_{n,d}(\mu)$ is a measure on the tensor space $(\mathbb{R}^n)^{\otimes p}$, which we identify with $\mathbb{R}^{n \cdot p}$, through the basis,

$$\{e_{i_1} \otimes \cdots \otimes e_{i_p} | 1 \leq i_1, \ldots, i_p \leq n\}.$$

For simplicity we will focus on the sub-space of 'principal' tensors, with basis,

$$\{e_{i_1} \otimes \cdots \otimes e_{i_p} | 1 \leq i_1 < \cdots < i_p \leq n\}.$$

The projection of $\mathcal{W}^p_{n,d}(\mu)$ will be denoted by $\widetilde{\mathcal{W}}^p_{n,d}(\mu)$.

Technicalities

$\mathcal{W}^p_{n,d}(\mu)$ is a measure on the tensor space $(\mathbb{R}^n)^{\otimes p}$, which we identify with $\mathbb{R}^{n \cdot p}$, through the basis,

$$\{e_{i_1} \otimes \cdots \otimes e_{i_p} | 1 \leq i_1, \ldots, i_p \leq n\}.$$

For simplicity we will focus on the sub-space of 'principal' tensors, with basis,

$$\{e_{i_1} \otimes \cdots \otimes e_{i_p} | 1 \leq i_1 < \cdots < i_p \leq n\}.$$

The projection of $\mathcal{W}^p_{n,d}(\mu)$ will be denoted by $\widetilde{\mathcal{W}}^p_{n,d}(\mu)$.

Technicalities

$\mathcal{W}^p_{n,d}(\mu)$ is a measure on the tensor space $(\mathbb{R}^n)^{\otimes p}$, which we identify with $\mathbb{R}^{n \cdot p}$, through the basis,

$$\{e_{i_1} \otimes \cdots \otimes e_{i_p} | 1 \leq i_1, \ldots, i_p \leq n\}.$$

For simplicity we will focus on the sub-space of 'principal' tensors, with basis,

$$\{e_{i_1} \otimes \cdots \otimes e_{i_p} | 1 \leq i_1 < \cdots < i_p \leq n\}.$$

The projection of $\mathcal{W}^p_{n,d}(\mu)$ will be denoted by $\widetilde{\mathcal{W}}^p_{n,d}(\mu)$.

Wishart Matrices

When $p = 2$ and $X \sim \mu$ is isotropic, $\mathcal{W}^2_{n,d}(\mu)$ can be realized as the law of

$$\frac{\mathbb{X}\mathbb{X}^T - d \cdot \mathrm{Id}}{\sqrt{d}}.$$

Here, $\mathbb{X}$ is an $n \times d$ matrix, with columns being *i.i.d.* copies of X.

In this case, $\widetilde{\mathcal{W}}^2_{n,d}(\mu)$ is the law of the upper triangular part.

Some Observations

Let us restrict our attention to the case $p = 2$.

- for fixed n, by the central limit theorem $\mathcal{W}^2_{n,d}(\mu) \to \mathcal{N}(0, \Sigma)$.
- If $n = d$, then the spectral measure of $\mathbb{X}\mathbb{X}^T$ converges to the Marchenko-Pastur distribution. In particular, $\mathcal{W}^2_{n,d}(\mu)$ is not Gaussian.

Question

How should n depend on d so that $\mathcal{W}^p_{n,d}(\mu)$ is approximately Gaussian.

Some Observations

Let us restrict our attention to the case $p = 2$.

- for fixed n, by the central limit theorem $\mathcal{W}^2_{n,d}(\mu) \to \mathcal{N}(0, \Sigma)$.
- If $n = d$, then the spectral measure of $\mathbb{X}\mathbb{X}^T$ converges to the Marchenko-Pastur distribution. In particular, $\mathcal{W}^2_{n,d}(\mu)$ is not Gaussian.

Question

How should n depend on d so that $\mathcal{W}^p_{n,d}(\mu)$ is approximately Gaussian.

Some Observations

Let us restrict our attention to the case $p = 2$.

- for fixed n, by the central limit theorem $\mathcal{W}^2_{n,d}(\mu) \to \mathcal{N}(0, \Sigma)$.
- If $n = d$, then the spectral measure of $\mathbb{X}\mathbb{X}^T$ converges to the Marchenko-Pastur distribution. In particular, $\mathcal{W}^2_{n,d}(\mu)$ is not Gaussian.

Question

How should n depend on d so that $\mathcal{W}^p_{n,d}(\mu)$ is approximately Gaussian.

Some Observations

Let us restrict our attention to the case $p = 2$.

- for fixed n, by the central limit theorem $\mathcal{W}_{n,d}^2(\mu) \to \mathcal{N}(0, \Sigma)$.
- If $n = d$, then the spectral measure of $\mathbb{X}\mathbb{X}^T$ converges to the Marchenko-Pastur distribution. In particular, $\mathcal{W}_{n,d}^2(\mu)$ is not Gaussian.

Question

How should n depend on d so that $\mathcal{W}_{n,d}^p(\mu)$ is approximately Gaussian.

Random Geometric Graphs

From now on, let γ stand for the standard Gaussian, in different dimensions. In (Bubeck, Ding, Eldan, Rácz 15') and independently in (Jiang, Li 15') it was shown,

- If $\frac{n^3}{d} \to 0$, then $\mathrm{TV}\left(\widetilde{\mathcal{W}}^2_{n,d}(\gamma), \gamma\right) \to 0$.

This is tight, in the sense,

- If $\frac{n^3}{d} \to \infty$, then $\mathrm{TV}\left(\widetilde{\mathcal{W}}^2_{n,d}(\gamma), \gamma\right) \to 1$.

(Rácz, Richey 16') shows that the phase transition is smooth.

Extensions

(Bubeck, Ganguly 15') extended the result to any log-concave product measure. That is, $\mathbb{X}_{i,j}$ are *i.i.d.* as $e^{-\varphi(x)}dx$ for some convex φ.

- Original motivation came from random geomteric graphs.
- (Fang, Koike 20') removed the log-concavity assumption.

Extensions

(Nourdin, Zheng 18') gave the following results, as an answer to questions raised in (Bubeck, Ganguly 15')

- If the rows of $\mathbb{X}$ are *i.i.d.* $\mathcal{N}(0, \Sigma)$, for some positive definite Σ. Then
$$W_1\left(\widetilde{\mathcal{W}}^2_{n,d}, \gamma\right) \lesssim \sqrt{\frac{n^3}{d}}.$$
(See also (Eldan, M 16'))
- $W_1\left(\widetilde{\mathcal{W}}^p_{n,d}(\gamma), \gamma\right) \lesssim \sqrt{\frac{n^{2p-1}}{d}}$.

Extensions

(Nourdin, Zheng 18') gave the following results, as an answer to questions raised in (Bubeck, Ganguly 15')

- If the rows of $\mathbb{X}$ are *i.i.d.* $\mathcal{N}(0, \Sigma)$, for some positive definite Σ. Then
$$W_1\left(\widetilde{\mathcal{W}}^2_{n,d}, \gamma\right) \lesssim \sqrt{\frac{n^3}{d}}.$$
(See also (Eldan, M 16'))
- $W_1\left(\widetilde{\mathcal{W}}^p_{n,d}(\gamma), \gamma\right) \lesssim \sqrt{\frac{n^{2p-1}}{d}}$.

Extensions

(Nourdin, Zheng 18') gave the following results, as an answer to questions raised in (Bubeck, Ganguly 15')

- If the rows of $\mathbb{X}$ are *i.i.d.* $\mathcal{N}(0, \Sigma)$, for some positive definite Σ. Then
$$W_1\left(\widetilde{\mathcal{W}}^2_{n,d}, \gamma\right) \lesssim \sqrt{\frac{n^3}{d}}.$$
(See also (Eldan, M 16'))
- $W_1\left(\widetilde{\mathcal{W}}^p_{n,d}(\gamma), \gamma\right) \lesssim \sqrt{\frac{n^{2p-1}}{d}}$.

Main Result

Today:

Theorem

If μ is a measure on $\mathbb{R}^n$ which is uniformly log-concave and unconditional, then

$$\mathrm{dist}\left(\widetilde{W}_{n,d}^{p}(\mu), \gamma\right) \lesssim \sqrt{\frac{n^{2p-1}}{d}}.$$

- dist stands from some notion of distance to be introduced soon. But could be replaced with W_2.
- The assumptions of uniform log-concavity and unconditionality may be relaxed.
- The result also holds for a large class of product measures.

Main Result

Today:

Theorem

If μ is a measure on $\mathbb{R}^n$ which is uniformly log-concave and unconditional, then

$$\mathrm{dist}\left(\widetilde{W}^p_{n,d}(\mu), \gamma\right) \lesssim \sqrt{\frac{n^{2p-1}}{d}}.$$

- dist stands from some notion of distance to be introduced soon. But could be replaced with W_2.
- The assumptions of uniform log-concavity and unconditionality may be relaxed.
- The result also holds for a large class of product measures.

Main Result

Today:

Theorem

If μ is a measure on $\mathbb{R}^n$ which is uniformly log-concave and unconditional, then

$$\mathrm{dist}\left(\widetilde{W}^p_{n,d}(\mu), \gamma\right) \lesssim \sqrt{\frac{n^{2p-1}}{d}}.$$

- dist stands from some notion of distance to be introduced soon. But could be replaced with W_2.
- The assumptions of uniform log-concavity and unconditionality may be relaxed.
- The result also holds for a large class of product measures.

Main Result

Today:

Theorem

If μ is a measure on $\mathbb{R}^n$ which is uniformly log-concave and unconditional, then

$$\mathrm{dist}\left(\widetilde{W}^{p}_{n,d}(\mu), \gamma\right) \lesssim \sqrt{\frac{n^{2p-1}}{d}}.$$

- dist stands from some notion of distance to be introduced soon. But could be replaced with W_2.
- The assumptions of uniform log-concavity and unconditionality may be relaxed.
- The result also holds for a large class of product measures.

The Challenge

By considering, $\frac{1}{\sqrt{d}}\sum_{i=1}^{d}\left(X_i^{\otimes p} - \mathbb{E}\left[X_i^{\otimes p}\right]\right)$, one may hope to be able to apply an estimate of the high-dimensional central limit theorem.

Optimistically, such estimates give:

$$\mathrm{dist}\left(\widetilde{W}^p_{n,d}(\mu), \gamma\right) \leq \frac{\mathbb{E}\left[\|X^{\otimes p}\|^3\right]}{\sqrt{d}}.$$

Thus, to obtain optimal convergence rates, we need to exploit the low dimensional structure of $\widetilde{W}^p_{n,d}(\mu)$.

The Challenge

By considering, $\frac{1}{\sqrt{d}}\sum_{i=1}^{d}\left(X_i^{\otimes p}-\mathbb{E}\left[X_i^{\otimes p}\right]\right)$, one may hope to be able to apply an estimate of the high-dimensional central limit theorem.

Optimistically, such estimates give:

$$\mathrm{dist}\left(\widetilde{W}^p_{n,d}(\mu),\gamma\right)\leq\frac{\mathbb{E}\left[\|X^{\otimes p}\|^3\right]}{\sqrt{d}}.$$

Thus, to obtain optimal convergence rates, we need to exploit the low dimensional structure of $\widetilde{W}^p_{n,d}(\mu)$.

The Challenge

By considering, $\frac{1}{\sqrt{d}} \sum_{i=1}^{d} \left(X_i^{\otimes p} - \mathbb{E}\left[X_i^{\otimes p} \right] \right)$, one may hope to be able to apply an estimate of the high-dimensional central limit theorem.

Optimistically, such estimates give:

$$\mathrm{dist}\left(\widetilde{W}^p_{n,d}(\mu), \gamma \right) \leq \frac{\mathbb{E}\left[\|X^{\otimes p}\|^3 \right]}{\sqrt{d}}.$$

Thus, to obtain optimal convergence rates, we need to exploit the low dimensional structure of $\widetilde{W}^p_{n,d}(\mu)$.

Stein's Method

Basic observation: If $G \sim \gamma$ on $\mathbb{R}^n$. Then, for any smooth test function $f : \mathbb{R}^n \to \mathbb{R}^n$,

$$\mathbb{E}\left[\langle G, f(G)\rangle\right] = \mathbb{E}\left[\mathrm{div} f(G)\right].$$

Moreover, the Gaussian is the only measure which satisfies this relation.

Stein's idea:

$$\mathbb{E}\left[\langle X, f(X)\rangle\right] \simeq \mathbb{E}\left[\mathrm{div} f(X)\right] \implies X \simeq G.$$

Stein's Method

Basic observation: If $G \sim \gamma$ on $\mathbb{R}^n$. Then, for any smooth test function $f : \mathbb{R}^n \to \mathbb{R}^n$,

$$\mathbb{E}\left[\langle G, f(G)\rangle\right] = \mathbb{E}\left[\mathrm{div} f(G)\right].$$

Moreover, the Gaussian is the only measure which satisfies this relation.

Stein's idea:

$$\mathbb{E}\left[\langle X, f(X)\rangle\right] \simeq \mathbb{E}\left[\mathrm{div} f(X)\right] \implies X \simeq G.$$

Stein Kernels

A Stein kernel of $X \sim \mu$ is a matrix valued map $\tau : \mathbb{R}^n \to M_n(\mathbb{R})$, such that

$$\mathbb{E}\left[\langle X, f(X)\rangle\right] = \mathbb{E}\left[\langle \tau(X), Df(X)\rangle_{HS}\right].$$

We have that $\tau \equiv \mathrm{Id}$ iff $\mu = \gamma$. The discrepancy is then defined as

$$S^2(\mu) = \mathbb{E}_\mu\left[\|\tau - \mathrm{Id}\|_{HS}^2\right].$$

Stein Kernels

A Stein kernel of $X \sim \mu$ is a matrix valued map $\tau : \mathbb{R}^n \to M_n(\mathbb{R})$, such that

$$\mathbb{E}\left[\langle X, f(X)\rangle\right] = \mathbb{E}\left[\langle \tau(X), Df(X)\rangle_{HS}\right].$$

We have that $\tau \equiv \mathrm{Id}$ iff $\mu = \gamma$. The discrepancy is then defined as

$$S^2(\mu) = \mathbb{E}_\mu\left[\|\tau - \mathrm{Id}\|_{HS}^2\right].$$

Stein Kernels - Properties

Stein kernels are well behaved under linear transformations. If τ_X is a stein kernel for X, and A is a linear transformation. Then

$$\tau_{AX}(x) := A\mathbb{E}\left[\tau_X(X)|AX = x\right]A^T,$$

is a Stein kernel for AX.

In particular, if $S_d := \frac{1}{\sqrt{d}} \sum_{i=1}^{d} X_i$,

$$\tau_{S_d}(x) = \frac{1}{d}\sum_{i=1}^{d}\mathbb{E}\left[\tau_X(X_i)|S_d = x\right],$$

is a Stein kernel for S_d.

Stein Kernels - Properties

Stein kernels are well behaved under linear transformations. If τ_X is a stein kernel for X, and A is a linear transformation. Then

$$\tau_{AX}(x) := A\mathbb{E}\left[\tau_X(X)|AX = x\right]A^T,$$

is a Stein kernel for AX.

In particular, if $S_d := \frac{1}{\sqrt{d}}\sum_{i=1}^{d} X_i$,

$$\tau_{S_d}(x) = \frac{1}{d}\sum_{i=1}^{d}\mathbb{E}\left[\tau_X(X_i)|S_d = x\right],$$

is a Stein kernel for S_d.

Stein's Discrepancy Along the CLT

If X is isotropic and $f(x) := x_i e_j$, we get

$$\delta_{i,j} = \mathbb{E}\left[\langle X, f(X)\rangle\right] = \mathbb{E}\left[\langle \tau_X(X), Df(X)\rangle\right] = \mathbb{E}\left[\tau_X(X)_{i,j}\right].$$

So, $\mathbb{E}\left[\tau_X(X)\right] = \mathrm{Id}$.

Thus,

$$S^2(S_d) = \mathbb{E}\left[||\tau_{S_d}(S_d) - \mathrm{Id}||_{HS}^2\right] = \mathbb{E}\left[\left|\left|\frac{1}{d}\sum_{i=1}^{d}\mathbb{E}\left[\tau_X(X_i) - \mathrm{Id}|S_d\right]\right|\right|_{HS}^2\right]$$

$$\leq \frac{1}{d^2}\left|\left|\mathbb{E}\left[\sum_{i=1}^{d}\tau_X(X_i) - \mathrm{Id}\right]\right|\right|_{HS}^2 = \frac{1}{d}\mathbb{E}\left[||\tau_X(X) - \mathrm{Id}||_{HS}^2\right]$$

$$= \frac{S^2(X)}{d}.$$

Stein's Discrepancy Along the CLT

If X is isotropic and $f(x) := x_i e_j$, we get

$$\delta_{i,j} = \mathbb{E}\left[\langle X, f(X)\rangle\right] = \mathbb{E}\left[\langle \tau_X(X), Df(X)\rangle\right] = \mathbb{E}\left[\tau_X(X)_{i,j}\right].$$

So, $\mathbb{E}\left[\tau_X(X)\right] = \mathrm{Id}$.

Thus,

$$\begin{aligned} S^2(S_d) &= \mathbb{E}\left[||\tau_{S_d}(S_d) - \mathrm{Id}||_{HS}^2\right] = \mathbb{E}\left[\left|\left|\frac{1}{d}\sum_{i=1}^{d}\mathbb{E}\left[\tau_X(X_i) - \mathrm{Id}|S_d\right]\right|\right|_{HS}^2\right] \\ &\leq \frac{1}{d^2}\left|\left|\mathbb{E}\left[\sum_{i=1}^{d}\tau_X(X_i) - \mathrm{Id}\right]\right|\right|_{HS}^2 = \frac{1}{d}\mathbb{E}\left[||\tau_X(X) - \mathrm{Id}||_{HS}^2\right] \\ &= \frac{S^2(X)}{d}. \end{aligned}$$

Stein's Discrepancy Compared to Other Distances

It's a nice exercise to show,

$$W_1(\mu, \gamma) \leq S(\mu).$$

What is more impressive is that

$$W_2(\mu, \gamma) \leq S(\mu),$$

as well, as shown in (Ledoux, Nourdin, Pecatti 14').

In fact,

$$\mathrm{Ent}(\mu||\gamma) \leq \frac{1}{2} S^2(\mu) \ln\left(1 + \frac{\mathrm{I}(\mu||\gamma)}{S^2(\mu)}\right).$$

Stein's Discrepancy Compared to Other Distances

It's a nice exercise to show,

$$W_1(\mu, \gamma) \leq S(\mu).$$

What is more impressive is that

$$W_2(\mu, \gamma) \leq S(\mu),$$

as well, as shown in (Ledoux, Nourdin, Pecatti 14').

In fact,

$$\mathrm{Ent}(\mu||\gamma) \leq \frac{1}{2} S^2(\mu) \ln \left(1 + \frac{\mathrm{I}(\mu||\gamma)}{S^2(\mu)}\right).$$

Stein's Discrepancy Compared to Other Distances

It's a nice exercise to show,

$$W_1(\mu, \gamma) \leq S(\mu).$$

What is more impressive is that

$$W_2(\mu, \gamma) \leq S(\mu),$$

as well, as shown in (Ledoux, Nourdin, Pecatti 14').
In fact,

$$\mathrm{Ent}(\mu || \gamma) \leq \frac{1}{2} S^2(\mu) \ln \left(1 + \frac{\mathrm{I}(\mu || \gamma)}{S^2(\mu)} \right).$$

Proof of Main Theorem

The main theorem is implied by

Lemma (Rank 1 Lemma)

Let $X \sim \mu$ be an isotropic random vector in $\mathbb{R}^n$. Then, for any transport map, such that $\varphi_* \gamma = \mu$, there exists a Stein kernel τ, such that

$$\mathbb{E}\left[\left\|\tau\left(X^{\otimes p} - \mathbb{E}\left[X^{\otimes p}\right]\right)\right\|_{HS}^2\right] \leq p^4 n \sqrt{\mathbb{E}\left[\|X\|^{8(p-1)}\right]\mathbb{E}\left[\|D\varphi(G)\|_{op}^8\right]}.$$

Proof of Main Theorem.

Let A be the linear projection, such that $A_* W^p_{n,d}(\mu) = \widetilde{W}^p_{n,d}(\mu)$. Take φ, with $D\varphi < L$, almost surely. Then

$$\begin{aligned}
S^2(\widetilde{W}^p_{n,d}(\mu)) &\leq \frac{S^2\left(A\left(X^{\otimes p} - \mathbb{E}\left[X^{\otimes p}\right]\right)\right)}{d} \\
&\leq \frac{C}{d}\left(\mathbb{E}\left[\|\tau\left(X^{\otimes p} - \mathbb{E}\left[X^{\otimes p}\right]\right)\|_{HS}^2\right] + \mathbb{E}\left[\|\mathrm{Id}\|_{HS}^2\right]\right) \\
&\leq \frac{C}{d}\left(\sqrt{\mathbb{E}\left[\|X\|^{8(p-1)}\right]\mathbb{E}\left[\|D\varphi(G)\|_{op}^8\right]} + n^p\right) \\
&\leq C\frac{n^{2p-1}}{d}.
\end{aligned}$$

□

Plan

The plan for the rest of the talk is to prove the rank 1 lemma. We need the following ingredients:

- Given a transport map ψ such that $\psi_* \gamma = \nu$. Construct a Stein kernel for ν with small norm.
- Show that if φ is such that $\varphi_* \gamma = \mu$ has tame tails, then this is also true for that map $x \to \varphi(x)^{\otimes p}$.
- Use the fact that $x \to \varphi(x)^{\otimes p}$ is a map from a low-dimensional space.

Analysis in Finite Dimensional Gauss Space

Let us first show how to construct a Stein kernel from a given transport map:

- We work in the space $L^2(\gamma)$. Introduce D as the total (weak) derivative operator and δ as its adjoint.
- The Orenstein-Uhlenbeck operator is defined as $L := -\delta \circ D$.
- Fact:
  there exists an operator L^{-1} such that for any f with $\mathbb{E}_\gamma[f] = 0$, $LL^{-1}f = f$.

Analysis in Finite Dimensional Gauss Space

Let us first show how to construct a Stein kernel from a given transport map:

- We work in the space $L^2(\gamma)$. Introduce D as the total (weak) derivative operator and δ as its adjoint.
- The Orenstein-Uhlenbeck operator is defined as $L := -\delta \circ D$.
- Fact:
  there exists an operator L^{-1} such that for any f with $\mathbb{E}_\gamma[f] = 0$, $LL^{-1}f = f$.

Analysis in Finite Dimensional Gauss Space

Let us first show how to construct a Stein kernel from a given transport map:

- We work in the space $L^2(\gamma)$. Introduce D as the total (weak) derivative operator and δ as its adjoint.
- The Orenstein-Uhlenbeck operator is defined as $L := -\delta \circ D$.
- Fact:
  there exists an operator L^{-1} such that for any f with $\mathbb{E}_\gamma[f] = 0$, $LL^{-1}f = f$.

Analysis in Finite Dimensional Gauss Space

Let us first show how to construct a Stein kernel from a given transport map:

- We work in the space $L^2(\gamma)$. Introduce D as the total (weak) derivative operator and δ as its adjoint.
- The Orenstein-Uhlenbeck operator is defined as $L := -\delta \circ D$.
- Fact:
  there exists an operator L^{-1} such that for any f with $\mathbb{E}_\gamma[f] = 0$, $LL^{-1}f = f$.

Constructing a Kernel

Lemma

Let γ_m be the standard Gaussian measure on $\mathbb{R}^m$ and let $\varphi : \mathbb{R}^m \to \mathbb{R}^N$. Set $\nu = \varphi_* \gamma_m$ and suppose that $\int\limits_{\mathbb{R}^N} x d\nu = 0$.

Then

$$\tau_\varphi(x) := \mathbb{E}_{G \sim \gamma_m}\left[(-DL^{-1})\varphi(G)(D\varphi(G))^T | \varphi(G) = x\right],$$

is a Stein kernel of ν.

Proof of Construction

Proof.

$$
\begin{aligned}
&\mathbb{E}\left[\langle Df(Y), \tau_\varphi(Y)\rangle_{HS}\right] \\
&= \mathbb{E}\left[\langle Df(Y), \mathbb{E}\left[(-DL^{-1})\varphi(G)(D\varphi(G))^T | \varphi(G) = Y\right]\rangle_{HS}\right] \\
&= \mathbb{E}\left[\langle Df(\varphi(G))D\varphi(G), (-DL^{-1})\varphi(G)\rangle_{HS}\right] \\
&= \mathbb{E}\left[\langle Df(\varphi(G)), (-DL^{-1})\varphi(G)\rangle_{HS}\right] \quad \text{(Chain rule)} \\
&= \mathbb{E}\left[\langle f \circ \varphi(G), (-\delta DL^{-1})\varphi(G)\rangle\right] \quad \text{(Adjoint operator)} \\
&= \mathbb{E}\left[\langle f \circ \varphi(G), LL^{-1}\varphi(G)\rangle\right] \quad L = -\delta D \\
&= \mathbb{E}\left[\langle f \circ \varphi(G), \varphi(G)\rangle\right] \quad \mathbb{E}[\varphi(G)] = 0 \\
&= \mathbb{E}\left[\langle f(Y), Y\rangle\right]. \quad \varphi_* \gamma_m = \nu
\end{aligned}
$$

□

Contraction

We have for any matrix norm, the following contraction property

$$\|\tau_\varphi(x)\|^2 \leq \mathbb{E}_{G\sim\gamma_m}\left[\left\|(-DL^{-1})\varphi(G)(D\varphi(G))^T\right\|^2 | \varphi(G) = x\right]$$
$$\leq \mathbb{E}_{G\sim\gamma_m}\left[\|D\varphi(G)\|^4 | \varphi(G) = x\right].$$

Thus, for example, if φ is 1-Lipschitz, $\|\tau_\varphi\|_{op} \leq 1$.

Contraction

We have for any matrix norm, the following contraction property

$$\|\tau_\varphi(x)\|^2 \leq \mathbb{E}_{G\sim\gamma_m}\left[\left\|(-DL^{-1})\varphi(G)(D\varphi(G))^T\right\|^2 |\varphi(G)=x\right]$$
$$\leq \mathbb{E}_{G\sim\gamma_m}\left[\|D\varphi(G)\|^4 |\varphi(G)=x\right].$$

Thus, for example, if φ is 1-Lipschitz, $\|\tau_\varphi\|_{op} \leq 1$.

Contraction

The contraction property can be obtained from the commutation relation

$$-DL^{-1}\varphi = \int_0^\infty e^{-t} P_t D\varphi dt,$$

where P_t is the Ornstein-Uhlenbeck semi-group.

For then

$$\tau_\varphi(x) = \int_0^\infty e^{-t} \mathbb{E}_{G\sim\gamma_m}\left[D\varphi(G) P_t\left(D\varphi(G)\right) | \varphi(G) = x\right].$$

Contraction

The contraction property can be obtained from the commutation relation

$$-DL^{-1}\varphi = \int_0^\infty e^{-t} P_t D\varphi dt,$$

where P_t is the Ornstein-Uhlenbeck semi-group.

For then

$$\tau_\varphi(x) = \int_0^\infty e^{-t} \mathbb{E}_{G\sim\gamma_m}\left[D\varphi(G) P_t\left(D\varphi(G)\right)|\varphi(G) = x\right].$$

Back to Rank 1 Tensors

Suppose we have a transport map, such that $\varphi_*\gamma = \mu$ and $X \sim \mu$. We now consider the map $u \to \varphi(u)^{\otimes p} - \mathbb{E}\left[X^{\otimes p}\right]$. Define

$$\begin{aligned}\tau(\tilde{v}^{\otimes p}) :=& \mathbb{E}\left[(-DL^{-1})\varphi(G)^{\otimes p}(D\varphi(G)^{\otimes p})^T|\varphi(G)^{\otimes p} = v^{\otimes p}\right] \\ =& \mathbb{E}\left[(-DL^{-1})\varphi(G)^{\otimes p}(D\varphi(G)^{\otimes p})^T|\varphi(G) = (\pm 1)^p v\right],\end{aligned}$$

which is a Stein kernel for $X^{\otimes p} - \mathbb{E}\left[X^{\otimes p}\right]$.

Back to Rank 1 Tensors

Recall, we wish to bound $\mathbb{E}\left[\|\tau(X^{\otimes p} - \mathbb{E}\left[X^{\otimes p}\right])\|_{HS}^2\right]$. For any two matrices A, B, we have

$$\|AB\|_{HS} \leq \mathrm{rank}(A)\|AB\|_{op}.$$

So, since $\mathrm{rank}(D\varphi(v)^{\otimes p}) \leq n$, contraction gives

$$\mathbb{E}\left[\|\tau(X^{\otimes p} - \mathbb{E}\left[X^{\otimes p}\right])\|_{HS}^2\right] \leq n\mathbb{E}\left[\|D\varphi(G)^{\otimes p}\|_{op}^4\right].$$

Back to Rank 1 Tensors

Recall, we wish to bound $\mathbb{E}\left[\|\tau(X^{\otimes p} - \mathbb{E}\left[X^{\otimes p}\right])\|_{HS}^2\right]$. For any two matrices A, B, we have

$$\|AB\|_{HS} \leq \mathrm{rank}(A)\|AB\|_{op}.$$

So, since $\mathrm{rank}(D\varphi(v)^{\otimes p}) \leq n$, contraction gives

$$\mathbb{E}\left[\|\tau(X^{\otimes p} - \mathbb{E}\left[X^{\otimes p}\right])\|_{HS}^2\right] \leq n\mathbb{E}\left[\|D\varphi(G)^{\otimes p}\|_{op}^4\right].$$

A Little Algebra

Write, for the Kronecker product,

$$D\varphi(v)^{\otimes p} = \sum_{i=1}^{p} \varphi(x)^{\otimes i-1} \otimes D\varphi(v) \otimes \varphi(v)^{\otimes p-i}.$$

This gives

$$\begin{aligned}\mathbb{E}\left[\|\tau(X^{\otimes p} - \mathbb{E}\left[X^{\otimes p}\right])\|_{HS}^2\right] &\leq np^4 \mathbb{E}\left[\|D\varphi(G)\|_{op}^4 \|\varphi(G)\|^{4(p-1)}\right] \\ &\leq np^4 \sqrt{\mathbb{E}\left[\|D\varphi(G)\|_{op}^8\right] \mathbb{E}\left[\|X\|^{8(p-1)}\right]}.\end{aligned}$$

A Little Algebra

Write, for the Kronecker product,

$$D\varphi(v)^{\otimes p} = \sum_{i=1}^{p} \varphi(x)^{\otimes i-1} \otimes D\varphi(v) \otimes \varphi(v)^{\otimes p-i}.$$

This gives

$$\begin{aligned}\mathbb{E}\left[\|\tau(X^{\otimes p} - \mathbb{E}\left[X^{\otimes p}\right])\|_{HS}^2\right] &\leq np^4 \mathbb{E}\left[\|D\varphi(G)\|_{op}^4 \|\varphi(G)\|^{4(p-1)}\right] \\ &\leq np^4 \sqrt{\mathbb{E}\left[\|D\varphi(G)\|_{op}^8\right] \mathbb{E}\left[\|X\|^{8(p-1)}\right]}.\end{aligned}$$

Future Directions

- What about the full tensor $\mathcal{W}^{p}_{n,d}(\mu)$? (Related to anti-concentration of polynomials)
- What about general log-concave measures (Related to the KLS and thin shell conjectures).
- What about other dependence structures?
- What about lower bounds when $p > 2$?

Future Directions

- What about the full tensor $\mathcal{W}^p_{n,d}(\mu)$? (Related to anti-concentration of polynomials)
- What about general log-concave measures (Related to the KLS and thin shell conjectures).
- What about other dependence structures?
- What about lower bounds when $p > 2$?

Future Directions

- What about the full tensor $\mathcal{W}^{p}_{n,d}(\mu)$? (Related to anti-concentration of polynomials)
- What about general log-concave measures (Related to the KLS and thin shell conjectures).
- What about other dependence structures?
- What about lower bounds when $p > 2$?

Future Directions

- What about the full tensor $\mathcal{W}^p_{n,d}(\mu)$? (Related to anti-concentration of polynomials)
- What about general log-concave measures (Related to the KLS and thin shell conjectures).
- What about other dependence structures?
- What about lower bounds when $p > 2$?

Future Directions

- What about the full tensor $\mathcal{W}^{p}_{n,d}(\mu)$? (Related to anti-concentration of polynomials)
- What about general log-concave measures (Related to the KLS and thin shell conjectures).
- What about other dependence structures?
- What about lower bounds when $p > 2$?

Thank you!