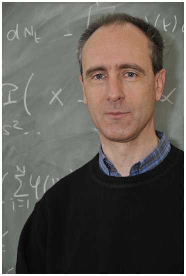
BEATING LOG-SOBOLEV: ONE STEIN'S KERNEL AT A TIME

Giovanni Peccati
(Luxembourg University)

Singapore: May 21, 2015

CREDITS

Based on two joint works: **(1)** Nourdin, Peccati and Swan (*JFA*, 2014), and **(2)** Ledoux, Nourdin and Peccati (*GAFA*, 2015).



A General Question

I shall use the notation $(\Omega, \mathscr{F}, \mathbf{P})$ for a generic probability space, with $\mathbf{E}$ indicating expectation with respect to $\mathbf{P}$. Fix two integers $k, d \geq 1$.

- ⋆ $G = (G_1, ..., G_k)$ is a k-dimensional vector of i.i.d. standard Gaussian random variables (the "underlying Gaussian field")
- ⋆ $F = (f_1(G), ..., f_d(G))$ is a d-dimensional vector of smooth (non-linear) transformations of G (the "unknown distribution") with identity covariance matrix. We assume that F has a density.
- ⋆ $N = (N_1, ..., N_d)$ is a d-dimensional vector of i.i.d. standard Gaussian random variables (the "target distribution").

Question: Can we (meaningfully!) bound the quantity

$$\mathbf{TV}(F, N) = \sup_{A \in \mathscr{B}(\mathbb{R}^d)} |\mathbf{P}(F \in A) - \mathbf{P}(N \in A)| \; ?$$

A GENERAL QUESTION

I shall use the notation $(\Omega, \mathscr{F}, \mathbf{P})$ for a generic probability space, with $\mathbf{E}$ indicating expectation with respect to $\mathbf{P}$. Fix two integers $k, d \geq 1$.

- $\star$ $G = (G_1, ..., G_k)$ is a k-dimensional vector of i.i.d. standard Gaussian random variables (the "underlying Gaussian field")
- $\star$ $F = (f_1(G), ..., f_d(G))$ is a d-dimensional vector of smooth (non-linear) transformations of G (the "unknown distribution") with identity covariance matrix. We assume that F has a density.
- $\star$ $N = (N_1, ..., N_d)$ is a d-dimensional vector of i.i.d. standard Gaussian random variables (the "target distribution").

Question: Can we (meaningfully!) bound the quantity

$$\mathbf{TV}(F, N) = \sup_{A \in \mathscr{B}(\mathbb{R}^d)} |\mathbf{P}(F \in A) - \mathbf{P}(N \in A)| \ ?$$

A General Question

I shall use the notation $(\Omega, \mathscr{F}, \mathbf{P})$ for a generic probability space, with $\mathbf{E}$ indicating expectation with respect to $\mathbf{P}$. Fix two integers $k, d \geq 1$.

- ⋆ $G = (G_1, ..., G_k)$ is a k-dimensional vector of i.i.d. standard Gaussian random variables (the "underlying Gaussian field")
- ⋆ $F = (f_1(G), ..., f_d(G))$ is a d-dimensional vector of smooth (non-linear) transformations of G (the "unknown distribution") with identity covariance matrix. We assume that F has a density.
- ⋆ $N = (N_1, ..., N_d)$ is a d-dimensional vector of i.i.d. standard Gaussian random variables (the "target distribution").

Question: Can we (meaningfully!) bound the quantity

$$\mathbf{TV}(F, N) = \sup_{A \in \mathscr{B}(\mathbb{R}^d)} |\mathbf{P}(F \in A) - \mathbf{P}(N \in A)| \text{ ?}$$

A General Question

I shall use the notation $(\Omega, \mathscr{F}, \mathbf{P})$ for a generic probability space, with $\mathbf{E}$ indicating expectation with respect to $\mathbf{P}$. Fix two integers $k, d \geq 1$.

- ⋆ $G = (G_1, ..., G_k)$ is a k-dimensional vector of i.i.d. standard Gaussian random variables (the "underlying Gaussian field")
- ⋆ $F = (f_1(G), ..., f_d(G))$ is a d-dimensional vector of smooth (non-linear) transformations of G (the "unknown distribution") with identity covariance matrix. We assume that F has a density.
- ⋆ $N = (N_1, ..., N_d)$ is a d-dimensional vector of i.i.d. standard Gaussian random variables (the "target distribution").

Question: Can we (meaningfully!) bound the quantity

$$\mathbf{TV}(F, N) = \sup_{A \in \mathscr{B}(\mathbb{R}^d)} |\mathbf{P}(F \in A) - \mathbf{P}(N \in A)| \ ?$$

A General Question

I shall use the notation $(\Omega, \mathscr{F}, \mathbf{P})$ for a generic probability space, with $\mathbf{E}$ indicating expectation with respect to $\mathbf{P}$. Fix two integers $k, d \geq 1$.

- ⋆ $G = (G_1, ..., G_k)$ is a k-dimensional vector of i.i.d. standard Gaussian random variables (the "underlying Gaussian field")
- ⋆ $F = (f_1(G), ..., f_d(G))$ is a d-dimensional vector of smooth (non-linear) transformations of G (the "unknown distribution") with identity covariance matrix. We assume that F has a density.
- ⋆ $N = (N_1, ..., N_d)$ is a d-dimensional vector of i.i.d. standard Gaussian random variables (the "target distribution").

Question: Can we (meaningfully!) bound the quantity

$$\mathbf{TV}(F, N) = \sup_{A \in \mathscr{B}(\mathbb{R}^d)} |\mathbf{P}(F \in A) - \mathbf{P}(N \in A)| \ ?$$

A General Question

I shall use the notation $(\Omega, \mathscr{F}, \mathbf{P})$ for a generic probability space, with $\mathbf{E}$ indicating expectation with respect to $\mathbf{P}$. Fix two integers $k, d \geq 1$.

- ⋆ $G = (G_1, ..., G_k)$ is a k-dimensional vector of i.i.d. standard Gaussian random variables (the "underlying Gaussian field")
- ⋆ $F = (f_1(G), ..., f_d(G))$ is a d-dimensional vector of smooth (non-linear) transformations of G (the "unknown distribution") with identity covariance matrix. We assume that F has a density.
- ⋆ $N = (N_1, ..., N_d)$ is a d-dimensional vector of i.i.d. standard Gaussian random variables (the "target distribution").

Question: Can we (meaningfully!) bound the quantity

$$\mathbf{TV}(F, N) = \sup_{A \in \mathscr{B}(\mathbb{R}^d)} |\mathbf{P}(F \in A) - \mathbf{P}(N \in A)| \ ?$$

NOTATION

The law of N is denoted by γ_d, that is

$$\begin{aligned}\gamma_d(\mathrm{d}x_1,...,\mathrm{d}x_d) &= \frac{1}{(2\pi)^{d/2}}\exp\left\{-\frac{1}{2}\sum_{i=1}^{d}x_i^2\right\}\mathrm{d}x_1\cdots\mathrm{d}x_d \\ &:= \phi(x)\mathrm{d}x.\end{aligned}$$

We shall assume that the law of F has a smooth density h with respect to γ_d, that is:

$$\mathbf{P}[F \in A] = \int_A h(x)\,\gamma_d(\mathrm{d}x), \quad A \subseteq \mathbb{R}^d.$$

This implies, in particular, that

$$\mathbf{TV}(F,N) = \frac{1}{2}\int_{\mathbb{R}^d} |h(x)-1|\,\gamma_d(\mathrm{d}x).$$

NOTATION

The law of N is denoted by γ_d, that is

$$\begin{aligned}\gamma_d(\mathrm{d}x_1, ..., \mathrm{d}x_d) &= \frac{1}{(2\pi)^{d/2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{d} x_i^2\right\} \mathrm{d}x_1 \cdots \mathrm{d}x_d \\ &:= \phi(x)\mathrm{d}x.\end{aligned}$$

We shall assume that the law of F has a smooth density h with respect to γ_d, that is:

$$\mathbf{P}[F \in A] = \int_A h(x)\,\gamma_d(\mathrm{d}x), \quad A \subseteq \mathbb{R}^d.$$

This implies, in particular, that

$$\mathbf{TV}(F, N) = \frac{1}{2}\int_{\mathbb{R}^d} |h(x) - 1|\,\gamma_d(\mathrm{d}x).$$

Notation

The law of N is denoted by γ_d, that is

$$\begin{aligned}\gamma_d(\mathrm{d}x_1, ..., \mathrm{d}x_d) &= \frac{1}{(2\pi)^{d/2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{d} x_i^2\right\} \mathrm{d}x_1 \cdots \mathrm{d}x_d \\ &:= \phi(x)\mathrm{d}x.\end{aligned}$$

We shall assume that the law of F has a smooth density h with respect to γ_d, that is:

$$\mathbf{P}[F \in A] = \int_A h(x)\, \gamma_d(\mathrm{d}x), \quad A \subseteq \mathbb{R}^d.$$

This implies, in particular, that

$$\mathbf{TV}(F, N) = \frac{1}{2}\int_{\mathbb{R}^d} |h(x) - 1|\, \gamma_d(\mathrm{d}x).$$

The Ornstein-Uhlenbeck semigroup

It will be useful to consider the standard interpolation:

$$F_t := e^{-t}F + \sqrt{1-e^{-2t}}N, \quad t \geq 0,$$

so that $F_0 = F$ and $F_\infty = N$.

It is easy to prove that the density of F_t is given by $P_t h$, where, for a given test function g,

$$P_t g(x) = \mathbf{E}[g(e^{-t}x + \sqrt{1-e^{-2t}}N)], \quad t \geq 0,$$

defines the **Ornstein-Uhlenbeck semigroup** (Mehler's form).

We denote by L the **generator** of $\{P_t\}$. It is well known that L (as an operator on $L^2(\gamma_k)$) has eigenvalues $0, -1, -2, ...,$; the corresponding eigenspaces $\{C_n : n \geq 0\}$ are the so-called **Wiener chaoses** associated with γ_k.

THE ORNSTEIN-UHLENBECK SEMIGROUP

It will be useful to consider the standard interpolation:

$$F_t := e^{-t}F + \sqrt{1 - e^{-2t}}N, \quad t \geq 0,$$

so that $F_0 = F$ and $F_\infty = N$.

It is easy to prove that the density of F_t is given by $P_t h$, where, for a given test function g,

$$P_t g(x) = \mathbf{E}[g(e^{-t}x + \sqrt{1 - e^{-2t}}N)], \quad t \geq 0,$$

defines the **Ornstein-Uhlenbeck semigroup** (Mehler's form).

We denote by L the **generator** of $\{P_t\}$. It is well known that L (as an operator on $L^2(\gamma_k)$) has eigenvalues $0, -1, -2, ...,$; the corresponding eigenspaces $\{C_n : n \geq 0\}$ are the so-called **Wiener chaoses** associated with γ_k.

The Ornstein-Uhlenbeck semigroup

It will be useful to consider the standard interpolation:

$$F_t := e^{-t}F + \sqrt{1 - e^{-2t}}N, \quad t \geq 0,$$

so that $F_0 = F$ and $F_\infty = N$.

It is easy to prove that the density of F_t is given by $P_t h$, where, for a given test function g,

$$P_t g(x) = \mathbf{E}[g(e^{-t}x + \sqrt{1 - e^{-2t}}N)], \quad t \geq 0,$$

defines the **Ornstein-Uhlenbeck semigroup** (Mehler's form).

We denote by L the **generator** of $\{P_t\}$. It is well known that L (as an operator on $L^2(\gamma_k)$) has eigenvalues $0, -1, -2, ...,$; the corresponding eigenspaces $\{C_n : n \geq 0\}$ are the so-called **Wiener chaoses** associated with γ_k.

The Stein Kernel

An application of integration by parts shows that, for every smooth mapping $g : \mathbb{R}^d \to \mathbb{R}$,

$$\mathbf{E}[Fg(F)] = \mathbf{E}[\tau_h \cdot \nabla g(F)] \quad \text{(as vectors)}$$

where τ_h is the $d \times d$ matrix given by

$$\tau_h^{i,j} = \tau_h^{i,j}(F) = \mathbf{E}\left[\langle \nabla f_j(G), -\nabla L^{-1} f_i(G) \rangle_{\mathbb{R}^k} \,|\, F \right].$$

(L^{-1} is the (pseudo)inverse of the OU semigroup on $L^2(\gamma_k)$).

We call τ_h the **Stein kernel** associated with F. Note that, if $h = 1$ (and therefore $F \overset{LAW}{=} N$), then necessarily $\tau_h =$ identity.

The Stein Kernel

An application of integration by parts shows that, for every smooth mapping $g : \mathbb{R}^d \to \mathbb{R}$,

$$\mathbf{E}[Fg(F)] = \mathbf{E}[\tau_h \cdot \nabla g(F)] \quad \text{(as vectors)}$$

where τ_h is the $d \times d$ matrix given by

$$\tau_h^{i,j} = \tau_h^{i,j}(F) = \mathbf{E}\left[\langle \nabla f_j(G), -\nabla L^{-1} f_i(G) \rangle_{\mathbb{R}^k} \,|\, F \right].$$

(L^{-1} is the (pseudo)inverse of the OU semigroup on $L^2(\gamma_k)$).

We call τ_h the **Stein kernel** associated with F. Note that, if $h = 1$ (and therefore $F \overset{LAW}{=} N$), then necessarily $\tau_h =$ identity.

THE STEIN KERNEL

An application of integration by parts shows that, for every smooth mapping $g : \mathbb{R}^d \to \mathbb{R}$,

$$\mathbf{E}[Fg(F)] = \mathbf{E}[\tau_h \cdot \nabla g(F)] \quad \text{(as vectors)}$$

where τ_h is the $d \times d$ matrix given by

$$\tau_h^{i,j} = \tau_h^{i,j}(F) = \mathbf{E}\left[\langle \nabla f_j(G), -\nabla L^{-1} f_i(G)\rangle_{\mathbb{R}^k} \,|\, F\right].$$

(L^{-1} is the (pseudo)inverse of the OU semigroup on $L^2(\gamma_k)$).

We call τ_h the **Stein kernel** associated with F. Note that, if $h = 1$ (and therefore $F \overset{LAW}{=} N$), then necessarily $\tau_h =$ identity.

Definition
The **Stein discrepancy** *S between F and N is defined as follows:*

$$S = S(F\|N) := \sqrt{\sum_{i,j=1}^{d} \mathbf{E}[(\tau_h^{i,j} - \delta_j^i)^2]}$$

A direct computation shows that

$$S(F_t\|N) \leq e^{-2t} S(F\|N), \quad t \geq 0.$$

Definition
The **Stein discrepancy** *S between F and N is defined as follows:*

$$S = S(F\|N) := \sqrt{\sum_{i,j=1}^{d} \mathbf{E}[(\tau_h^{i,j} - \delta_j^i)^2]}$$

A direct computation shows that

$$S(F_t\|N) \leq e^{-2t} S(F\|N), \quad t \geq 0.$$

Some estimates

Can one use S to rigorously measure the distance between F and N?

Stein's method (Stein '72, '86) allows one to obtain the following bounds (see Nourdin and Peccati, '09, '12; Nourdin and Rosinski, '12).

- $\star$ When $d = 1$,

$$\mathbf{TV}(F, N) \leq 2S(F \| N).$$

- $\star$ When $d = 1$ and F belongs to the qth Wiener chaos of γ_k ($q \geq 2$), then

$$\mathbf{TV}(F, N) \leq 2S(F \| N) \leq 2\sqrt{\frac{q-1}{3q}}\sqrt{E[F^4] - 3}$$

(note that $3 = \mathbf{E}[N^4]$). See Nualart and Peccati (2005).

SOME ESTIMATES

Can one use S to rigorously measure the distance between F and N?

Stein's method (Stein '72, '86) allows one to obtain the following bounds (see Nourdin and Peccati, '09, '12; Nourdin and Rosinski, '12).

- ⋆ When $d = 1$,

$$\mathbf{TV}(F,N) \leq 2S(F\|N).$$

- ⋆ When $d = 1$ and F belongs to the qth Wiener chaos of γ_k ($q \geq 2$), then

$$\mathbf{TV}(F,N) \leq 2S(F\|N) \leq 2\sqrt{\frac{q-1}{3q}}\sqrt{E[F^4]-3}$$

(note that $3 = \mathbf{E}[N^4]$). See Nualart and Peccati (2005).

Some estimates

Can one use S to rigorously measure the distance between F and N?

Stein's method (Stein '72, '86) allows one to obtain the following bounds (see Nourdin and Peccati, '09, '12; Nourdin and Rosinski, '12).

- ⋆ When $d = 1$,

$$\mathbf{TV}(F, N) \leq 2S(F\|N).$$

- ⋆ When $d = 1$ and F belongs to the qth Wiener chaos of γ_k ($q \geq 2$), then

$$\mathbf{TV}(F, N) \leq 2S(F\|N) \leq 2\sqrt{\frac{q-1}{3q}}\sqrt{E[F^4] - 3}$$

(note that $3 = \mathbf{E}[N^4]$). See Nualart and Peccati (2005).

Some estimates

Can one use S to rigorously measure the distance between F and N?

Stein's method (Stein '72, '86) allows one to obtain the following bounds (see Nourdin and Peccati, '09, '12; Nourdin and Rosinski, '12).

- $\star$ When $d = 1$,
$$\mathbf{TV}(F, N) \leq 2S(F\|N).$$

- $\star$ When $d = 1$ and F belongs to the qth Wiener chaos of γ_k ($q \geq 2$), then
$$\mathbf{TV}(F, N) \leq 2S(F\|N) \leq 2\sqrt{\frac{q-1}{3q}}\sqrt{E[F^4] - 3}$$
(note that $3 = \mathbf{E}[N^4]$). See Nualart and Peccati (2005).

SOME ESTIMATES

- ⋆ In general, for any $d \geq 1$,

$$\mathbf{Wass}_1(F,N) := \sup_{g \in \mathbf{Lip}(1)} |\mathbf{E}(g(F)) - \mathbf{E}(g(N))| \leq S(F\|N).$$

- ⋆ Finally, for any $d \geq 1$ and when each F_i is chaotic,

$$\mathbf{Wass}_1(F,N) \leq S(F\|N) \leq \sqrt{\mathbf{E}[\|F\|^4_{\mathbb{R}^d}] - \mathbf{E}[\|N\|^4_{\mathbb{R}^d}]}.$$

- ⋆ Very large scope of applications, e.g.: **(1)** *variations of Gaussian-subordinated processes*, **(2)** *local times of fractional processes*, **(3)** *zeros of random polynomials*, **(4)** *statistical analysis of spherical fields*, **(5)** *excursion sets of random fields on homogeneous spaces*, **(6)** *random matrices*, **(7)** *universality principles*.

SOME ESTIMATES

- ⋆ In general, for any $d \geq 1$,

$$\mathbf{Wass}_1(F,N) := \sup_{g \in \mathbf{Lip}(1)} |\mathbf{E}(g(F)) - \mathbf{E}(g(N))| \leq S(F\|N).$$

- ⋆ Finally, for any $d \geq 1$ and when each F_i is chaotic,

$$\mathbf{Wass}_1(F,N) \leq S(F\|N) \leq \sqrt{\mathbf{E}[\|F\|^4_{\mathbb{R}^d}] - \mathbf{E}[\|N\|^4_{\mathbb{R}^d}]}.$$

- ⋆ Very large scope of applications, e.g.: **(1)** *variations of Gaussian-subordinated processes*, **(2)** *local times of fractional processes*, **(3)** *zeros of random polynomials*, **(4)** *statistical analysis of spherical fields*, **(5)** *excursion sets of random fields on homogeneous spaces*, **(6)** *random matrices*, **(7)** *universality principles*.

Some estimates

- $\star$ In general, for any $d \geq 1$,

$$\mathbf{Wass}_1(F,N) := \sup_{g \in \mathbf{Lip}(1)} |\mathbf{E}(g(F)) - \mathbf{E}(g(N))| \leq S(F\|N).$$

- $\star$ Finally, for any $d \geq 1$ and when each F_i is chaotic,

$$\mathbf{Wass}_1(F,N) \leq S(F\|N) \leq \sqrt{\mathbf{E}[\|F\|^4_{\mathbb{R}^d}] - \mathbf{E}[\|N\|^4_{\mathbb{R}^d}]}.$$

- $\star$ Very large scope of applications, e.g.: **(1)** *variations of Gaussian-subordinated processes*, **(2)** *local times of fractional processes*, **(3)** *zeros of random polynomials*, **(4)** *statistical analysis of spherical fields*, **(5)** *excursion sets of random fields on homogeneous spaces*, **(6)** *random matrices*, **(7)** *universality principles*.

From Wasserstein to Total Variation

- ⋆ For a general dimension d, going from $\mathbf{Wass}_1$ to **TV** is a remarkably difficult task.
- ⋆ For instance, Nourdin, Nualart and Polly (2013), have proved that, if $F_n \overset{LAW}{\to} N$, and F_n lives in the sum of the first q chaoses

$$\mathbf{TV}(F_n, N) = O(1)\Big(\mathbf{E}[\|F_n\|^4_{\mathbb{R}^d}] - \mathbf{E}[\|N\|^4_{\mathbb{R}^d}]\Big)^{\alpha},$$

for every

$$\alpha < \frac{1}{1 + (d+1)(3 + 4d(q+1))}.$$

- ⋆ We will address this task by using the notions of **entropy** (H), **Fisher information** (I), and **2-Wasserstein distance** (W).

From Wasserstein to Total Variation

- ⋆ For a general dimension d, going from $\mathbf{Wass}_1$ to $\mathbf{TV}$ is a remarkably difficult task.
- ⋆ For instance, Nourdin, Nualart and Polly (2013), have proved that, if $F_n \overset{LAW}{\to} N$, and F_n lives in the sum of the first q chaoses

$$\mathbf{TV}(F_n, N) = O(1)\Big(\mathbf{E}[\|F_n\|^4_{\mathbb{R}^d}] - \mathbf{E}[\|N\|^4_{\mathbb{R}^d}]\Big)^{\alpha},$$

for every

$$\alpha < \frac{1}{1 + (d+1)(3 + 4d(q+1))}.$$

- ⋆ We will address this task by using the notions of **entropy** (H), **Fisher information** (I), and **2-Wasserstein distance** (W).

From Wasserstein to Total Variation

- ⋆ For a general dimension d, going from $\mathbf{Wass}_1$ to $\mathbf{TV}$ is a remarkably difficult task.
- ⋆ For instance, Nourdin, Nualart and Polly (2013), have proved that, if $F_n \stackrel{LAW}{\to} N$, and F_n lives in the sum of the first q chaoses

$$\mathbf{TV}(F_n, N) = O(1)\Big(\mathbf{E}[\|F_n\|^4_{\mathbb{R}^d}] - \mathbf{E}[\|N\|^4_{\mathbb{R}^d}]\Big)^{\alpha},$$

for every

$$\alpha < \frac{1}{1 + (d+1)(3 + 4d(q+1))}.$$

- ⋆ We will address this task by using the notions of **entropy** (H), **Fisher information** (I), and **2-Wasserstein distance** (W).

FROM WASSERSTEIN TO TOTAL VARIATION

- ⋆ For a general dimension d, going from $\mathbf{Wass}_1$ to $\mathbf{TV}$ is a remarkably difficult task.
- ⋆ For instance, Nourdin, Nualart and Polly (2013), have proved that, if $F_n \overset{LAW}{\to} N$, and F_n lives in the sum of the first q chaoses

$$\mathbf{TV}(F_n, N) = O(1)\Big(\mathbf{E}[\|F_n\|^4_{\mathbb{R}^d}] - \mathbf{E}[\|N\|^4_{\mathbb{R}^d}]\Big)^{\alpha},$$

for every

$$\alpha < \frac{1}{1 + (d+1)(3 + 4d(q+1))}.$$

- ⋆ We will address this task by using the notions of **entropy** (H), **Fisher information** (I), and **2-Wasserstein distance** (W).

Definition

The **relative entropy** *of F with respect to N is given by*

$$H(F\|N) := \int_{\mathbb{R}^d} h(x)\log h(x)\,\gamma_d(\mathrm{d}x)$$
$$= \mathbf{E}[-\log\phi(N)] - \mathbf{E}[-\log p(F)] = \mathbf{Ent}(N) - \mathbf{Ent}(F).$$

(ϕ = density of N; $p = h\phi$ = density of F)

Recall **Pinsker's inequality**:

$$\mathbf{TV}(F,N)^2 \leq \frac{1}{2}H(F\|N).$$

Definition

The **relative entropy** *of F with respect to N is given by*

$$H(F\|N) := \int_{\mathbb{R}^d} h(x)\log h(x)\,\gamma_d(\mathrm{d}x)$$
$$= \mathbf{E}[-\log\phi(N)] - \mathbf{E}[-\log p(F)] = \mathbf{Ent}(N) - \mathbf{Ent}(F).$$

(ϕ = density of N; $p = h\phi$ = density of F)

Recall **Pinsker's inequality**:

$$\mathbf{TV}(F,N)^2 \leq \frac{1}{2}H(F\|N).$$

Definition

The **relative Fisher information** *of F with respect to N is given by*

$$\begin{aligned} I(F\|N) &:= \int_{\mathbb{R}^d} \frac{|\nabla h(x)|^2}{h(x)}\, \gamma_d(\mathrm{d}x) = \mathbf{E}|\nabla \log h(F)|^2 \\ &= \mathbf{E}[|\nabla \log p(F) - \nabla \log \phi(F)|^2] \end{aligned}$$

(ϕ = density of N; $p = h\phi$ = density of F)

It is well-known that:

$$I(F_t\|N) \le e^{-2t} I(F\|N),\ t \ge 0 \text{ (exponential decay)}.$$

Definition

The **relative Fisher information** *of F with respect to N is given by*

$$\begin{aligned} I(F\|N) &:= \int_{\mathbb{R}^d} \frac{|\nabla h(x)|^2}{h(x)}\, \gamma_d(\mathrm{d}x) = \mathbf{E}|\nabla \log h(F)|^2 \\ &= \mathbf{E}[|\nabla \log p(F) - \nabla \log \phi(F)|^2] \end{aligned}$$

(ϕ = density of N; $p = h\phi$ = density of F)

It is well-known that:

$$I(F_t\|N) \le e^{-2t} I(F\|N),\ t \ge 0 \textbf{ (exponential decay)}.$$

Important relations

- ⋆ A famous formula, due to **de Bruijn**, states that

$$H(F\|N) = \int_0^\infty I(F_t\|N)\,\mathrm{d}t$$

(the proof is based on the use of the heat equation).

- ⋆ Using the exponential decay of $t \mapsto I(F_t\|N)$, we deduce immediately the **log-Sobolev inequality** (Gross, 1972)

$$H(F\|N) \le I(F\|N) \int_0^\infty e^{-2t}\,\mathrm{d}t = \frac{1}{2} I(F\|N).$$

Important relations

- ⋆ A famous formula, due to **de Bruijn**, states that

$$H(F\|N) = \int_0^\infty I(F_t\|N)\,\mathrm{d}t$$

(the proof is based on the use of the heat equation).

- ⋆ Using the exponential decay of $t \mapsto I(F_t\|N)$, we deduce immediately the **log-Sobolev inequality** (Gross, 1972)

$$H(F\|N) \le I(F\|N) \int_0^\infty e^{-2t}\,\mathrm{d}t = \frac{1}{2} I(F\|N).$$

2-Wasserstein distance

Definition

The **2-Wasserstein distance** *between the laws of F and N is given by*

$$W(F,N) = \inf \sqrt{\mathbf{E}\|X - Y\|_{\mathbb{R}^d}^2}$$

where the infimum runs over all pairs (X, Y) *such that* $X \overset{LAW}{=} F$ *and* $Y \overset{LAW}{=} N$.

SOME CLASSIC RELATIONS

- ⋆ The famous **Talagrand's transportation inequality** (Talagrand, 1996) states that

$$W(F,N) \leq \sqrt{2H(F\|N)}.$$

- ⋆ In Otto-Villani (2001), it is proved that

$$W(F,N) \leq \int_0^\infty \sqrt{I(F_t\|N)}\,\mathrm{d}t.$$

- ⋆ Finally, in Otto-Villani (2001) one can find the well-known **HWI inequality**

$$H(F\|N) \leq W(F,N)\sqrt{I(F\|N)} - \frac{1}{2}W^2(F,N).$$

SOME CLASSIC RELATIONS

- ⋆ The famous **Talagrand's transportation inequality** (Talagrand, 1996) states that

$$W(F,N) \leq \sqrt{2H(F\|N)}.$$

- ⋆ In Otto-Villani (2001), it is proved that

$$W(F,N) \leq \int_0^\infty \sqrt{I(F_t\|N)}\,\mathrm{d}t.$$

- ⋆ Finally, in Otto-Villani (2001) one can find the well-known **HWI inequality**

$$H(F\|N) \leq W(F,N)\sqrt{I(F\|N)} - \frac{1}{2}W^2(F,N).$$

SOME CLASSIC RELATIONS

- ⋆ The famous **Talagrand's transportation inequality** (Talagrand, 1996) states that

$$W(F,N) \leq \sqrt{2H(F\|N)}.$$

- ⋆ In Otto-Villani (2001), it is proved that

$$W(F,N) \leq \int_0^\infty \sqrt{I(F_t\|N)}\,\mathrm{d}t.$$

- ⋆ Finally, in Otto-Villani (2001) one can find the well-known **HWI inequality**

$$H(F\|N) \leq W(F,N)\sqrt{I(F\|N)} - \frac{1}{2}W^2(F,N).$$

Proposition (Ledoux, Nourdin, Peccati (2015))

Assume F admits a Stein's kernel τ_h:

$$\mathbf{E}[Fg(F)] = \mathbf{E}[\tau_h \cdot \nabla g(F)] \quad \textit{(g smooth).}$$

Then,

$$I(F_t \| N) \leq \frac{e^{-4t}}{1 - e^{-2t}} S^2(F \| N), \quad t > 0.$$

Proof: Writing p_t for the density of F_t, the estimate follows from the relation

$$\nabla \log p_t(F_t) - \nabla \log \phi(F_t) \quad = \quad -\frac{e^{-2t}}{\sqrt{1 - e^{-2t}}} \mathbf{E}[(\mathbf{Id} - \tau_h(F))N \,|\, F_t],$$

that can be easily verified by a direct computation.

THE ROLE OF STEIN DISCREPANCY

Proposition (Ledoux, Nourdin, Peccati (2015))

Assume F admits a Stein's kernel τ_h:

$$\mathbf{E}[Fg(F)] = \mathbf{E}[\tau_h \cdot \nabla g(F)] \quad \textit{(g smooth).}$$

Then,

$$I(F_t \| N) \leq \frac{e^{-4t}}{1 - e^{-2t}} S^2(F \| N), \quad t > 0.$$

Proof: Writing p_t for the density of F_t, the estimate follows from the relation

$$\nabla \log p_t(F_t) - \nabla \log \phi(F_t) = -\frac{e^{-2t}}{\sqrt{1 - e^{-2t}}} \mathbf{E}[(\mathbf{Id} - \tau_h(F))N \,|\, F_t],$$

that can be easily verified by a direct computation.

Proposition (Ledoux, Nourdin, Peccati (2015))

Assume F admits a Stein's kernel τ_h:

$$\mathbf{E}[Fg(F)] = \mathbf{E}[\tau_h \cdot \nabla g(F)] \quad \textit{(g smooth).}$$

Then,

$$I(F_t \| N) \leq \frac{e^{-4t}}{1 - e^{-2t}} S^2(F \| N), \quad t > 0.$$

Proof: Writing p_t for the density of F_t, the estimate follows from the relation

$$\nabla \log p_t(F_t) - \nabla \log \phi(F_t) \quad = \quad -\frac{e^{-2t}}{\sqrt{1 - e^{-2t}}} \mathbf{E}[(\mathbf{Id} - \tau_h(F))N \,|\, F_t],$$

that can be easily verified by a direct computation.

The HSI inequality

The following result is proved in Ledoux, Nourdin and Peccati (2015)

Theorem (HSI inequality)

One has that:

$$H(F\|N) \leq \frac{S^2(F\|N)}{2} \log\left(1 + \frac{I(F\|N)}{S^2(F\|N)}\right).$$

Remark: since (for $x, y > 0$)

$$x \log(1 + y/x) \leq y,$$

this estimate (strictly) improves the Log-Sobolev inequality $2H \leq I$.

The HSI inequality

The following result is proved in Ledoux, Nourdin and Peccati (2015)

Theorem (HSI inequality)

One has that:

$$H(F\|N) \leq \frac{S^2(F\|N)}{2} \log\left(1 + \frac{I(F\|N)}{S^2(F\|N)}\right).$$

Remark: since (for $x, y > 0$)

$$x \log(1 + y/x) \leq y,$$

this estimate (strictly) improves the Log-Sobolev inequality $2H \leq I$.

Remarks

⋆ The proof follows from the estimate

$$\begin{aligned} H(F\|N) &= \int_0^\infty I(F_t\|N)\mathrm{d}t \\ &= \int_0^x I(F_t\|N)\mathrm{d}t + \int_x^\infty I(F_t\|N)\mathrm{d}t \\ &\leq I(F\|N)\int_0^x e^{-2t}\mathrm{d}t + S^2(F\|N)\int_x^\infty \frac{e^{-4t}}{1-e^{-2t}}\mathrm{d}t, \end{aligned}$$

and then by optimising in x.

⋆ It is easy to cook up examples where the RHS of HSI converges to zero, while the RHS of HWI (and therefore of log-Sobolev) "explodes" to infinity. We were not able to construct examples going in the other direction.

REMARKS

⋆ The proof follows from the estimate

$$\begin{aligned} H(F\|N) &= \int_0^\infty I(F_t\|N)\mathrm{d}t \\ &= \int_0^x I(F_t\|N)\mathrm{d}t + \int_x^\infty I(F_t\|N)\mathrm{d}t \\ &\le I(F\|N)\int_0^x e^{-2t}\mathrm{d}t + S^2(F\|N)\int_x^\infty \frac{e^{-4t}}{1-e^{-2t}}\mathrm{d}t, \end{aligned}$$

and then by optimising in x.

⋆ It is easy to cook up examples where the RHS of HSI converges to zero, while the RHS of HWI (and therefore of log-Sobolev) "explodes" to infinity. We were not able to construct examples going in the other direction.

Remarks

- ⋆ The proof follows from the estimate

$$\begin{aligned} H(F\|N) &= \int_0^\infty I(F_t\|N)\mathrm{d}t \\ &= \int_0^x I(F_t\|N)\mathrm{d}t + \int_x^\infty I(F_t\|N)\mathrm{d}t \\ &\leq I(F\|N)\int_0^x e^{-2t}\mathrm{d}t + S^2(F\|N)\int_x^\infty \frac{e^{-4t}}{1-e^{-2t}}\mathrm{d}t, \end{aligned}$$

  and then by optimising in x.
- ⋆ It is easy to cook up examples where the RHS of HSI converges to zero, while the RHS of HWI (and therefore of log-Sobolev) "explodes" to infinity. We were not able to construct examples going in the other direction.

The HWS inequality

A suitable modification of the above approach gives the following bound (Ledoux, Nourdin, Peccati, 2015)

Theorem (HWS inequality)

Under the above notation,

$$W(F,N) \leq S(F\|N)\ \arccos\left(e^{-\frac{H(F\|N)}{S^2(F\|N)}}\right).$$

Remark: this estimate improves Talagrand's transportation inequality $W \leq \sqrt{2H}$. The reason is that

$$\arccos e^{-x} \leq \sqrt{2x},$$

for every $x \geq 0$.

A suitable modification of the above approach gives the following bound (Ledoux, Nourdin, Peccati, 2015)

Theorem (HWS inequality)

Under the above notation,

$$W(F,N) \leq S(F\|N) \; \arccos\left(e^{-\frac{H(F\|N)}{S^2(F\|N)}}\right).$$

Remark: this estimate improves Talagrand's transportation inequality $W \leq \sqrt{2H}$. The reason is that

$$\arccos e^{-x} \leq \sqrt{2x},$$

for every $x \geq 0$.

THE HWS INEQUALITY

A suitable modification of the above approach gives the following bound (Ledoux, Nourdin, Peccati, 2015)

Theorem (HWS inequality)

Under the above notation,

$$W(F,N) \leq S(F\|N)\ \arccos\left(e^{-\frac{H(F\|N)}{S^2(F\|N)}}\right).$$

Remark: this estimate improves Talagrand's transportation inequality $W \leq \sqrt{2H}$. The reason is that

$$\arccos e^{-x} \leq \sqrt{2x},$$

for every $x \geq 0$.

REMARKS

- ⋆ Although we have the hierarchy

$$\textbf{HSI}, \ \textbf{HWI} \Rightarrow \textbf{Log-Sobolev} \Rightarrow \textbf{Talagrand},$$

  the relations between HSI and HWI are far from being clear.
- ⋆ More to the point, when trying to deduce a 'HWSI' inequality by a similar optimisation procedure, one finds that the minimum is either HWI or HSI. This phenomenon is difficult to interpret at the moment.

REMARKS

- ⋆ Although we have the hierarchy

$$\textbf{HSI},\ \textbf{HWI} \Rightarrow \textbf{Log-Sobolev} \Rightarrow \textbf{Talagrand},$$

the relations between HSI and HWI are far from being clear.

- ⋆ More to the point, when trying to deduce a ‘HWSI’ inequality by a similar optimisation procedure, one finds that the minimum is either HWI or HSI. This phenomenon is difficult to interpret at the moment.

SOME APPLICATIONS & EXTENSIONS

⋆ Reinforced **convergence to equilibrium**:

$$H(F_t \,\|\, N) \leq \frac{e^{-4t}}{1-e^{-2t}} S^2(F\|N)$$

(compare with the estimate: $H(F_t \,\|\, N) \leq e^{-2t} H(F \,\|\, N)$).

⋆ **Concentration** via L^p norms of Lipschitz functions:

$$\mathbf{E}[|u(F)|^p]^{1/p} \leq C\left[S_p + \sqrt{p} + \sqrt{pS_p}\right],$$

where

$$S_p := \left(\mathbf{E}\|\mathbf{Id} - \tau_h\|_{HS}^p\right)^{1/p}.$$

⋆ Also,

$$W_p(F,N) \leq C_p\, S_p.$$

Some Applications & Extensions

⋆ Reinforced **convergence to equilibrium**:

$$H(F_t \,\|\, N) \le \frac{e^{-4t}}{1 - e^{-2t}} S^2(F\|N)$$

(compare with the estimate: $H(F_t \,\|\, N) \le e^{-2t} H(F \,\|\, N)$).

⋆ **Concentration** via L^p norms of Lipschitz functions:

$$\mathbf{E}[|u(F)|^p]^{1/p} \le C\left[S_p + \sqrt{p} + \sqrt{pS_p}\right],$$

where

$$S_p := \left(\mathbf{E}\|\mathbf{Id} - \tau_h\|_{HS}^p\right)^{1/p}.$$

⋆ Also,

$$W_p(F, N) \le C_p\, S_p.$$

SOME APPLICATIONS & EXTENSIONS

- ⋆ Reinforced **convergence to equilibrium**:

$$H(F_t \,\|\, N) \leq \frac{e^{-4t}}{1-e^{-2t}} S^2(F\|N)$$

(compare with the estimate: $H(F_t \,\|\, N) \leq e^{-2t} H(F \,\|\, N)$).

- ⋆ **Concentration** via L^p norms of Lipschitz functions:

$$\mathbf{E}[|u(F)|^p]^{1/p} \leq C\left[S_p + \sqrt{p} + \sqrt{pS_p}\right],$$

where

$$S_p := \left(\mathbf{E}\|\mathbf{Id} - \tau_h\|_{HS}^p\right)^{1/p}.$$

- ⋆ Also,

$$W_p(F, N) \leq C_p\, S_p.$$

EXTENSIONS

- ⋆ HSI can be suitably extended to the case where the target distribution μ is the invariant measure of a symmetric Markov semigroup $\{P_t\}$ with generator

$$\mathcal{L}f = \langle a, \mathrm{Hess} f\rangle_{HS} + b \cdot \nabla f.$$

- ⋆ In this case, we require the Stein kernel τ_ν to verify the relation

$$\int b \cdot \nabla f \, \mathrm{d}\nu + \int \langle \tau_\nu, \mathrm{Hess} f\rangle_{HS} \, \mathrm{d}\nu = 0$$

and therefore

$$S^2(\nu \,\|\, \mu) = \int \|a^{-1/2}\tau_\nu a^{-1/2} - \mathbf{Id}\|^2 \mathrm{d}\nu.$$

- ⋆ The conditions for HSI to hold involve iterated gradients Γ_n of orders $n = 1, 2, 3$.

EXTENSIONS

⋆ HSI can be suitably extended to the case where the target distribution μ is the invariant measure of a symmetric Markov semigroup $\{P_t\}$ with generator

$$\mathcal{L}f = \langle a, \mathrm{Hess} f\rangle_{HS} + b \cdot \nabla f.$$

⋆ In this case, we require the Stein kernel τ_ν to verify the relation

$$\int b \cdot \nabla f \, \mathrm{d}\nu + \int \langle \tau_\nu, \mathrm{Hess} f\rangle_{HS} \, \mathrm{d}\nu = 0$$

and therefore

$$S^2(\nu \,\|\, \mu) = \int \|a^{-1/2} \tau_\nu a^{-1/2} - \mathbf{Id}\|^2 \mathrm{d}\nu.$$

⋆ The conditions for HSI to hold involve iterated gradients Γ_n of orders $n = 1, 2, 3$.

EXTENSIONS

- ⋆ HSI can be suitably extended to the case where the target distribution μ is the invariant measure of a symmetric Markov semigroup $\{P_t\}$ with generator

$$\mathcal{L}f = \langle a, \mathrm{Hess} f\rangle_{HS} + b \cdot \nabla f.$$

- ⋆ In this case, we require the Stein kernel τ_ν to verify the relation

$$\int b \cdot \nabla f \, \mathrm{d}\nu + \int \langle \tau_\nu, \mathrm{Hess} f\rangle_{HS} \, \mathrm{d}\nu = 0$$

and therefore

$$S^2(\nu \,\|\, \mu) = \int \|a^{-1/2}\tau_\nu a^{-1/2} - \mathbf{Id}\|^2 \mathrm{d}\nu.$$

- ⋆ The conditions for HSI to hold involve iterated gradients Γ_n of orders $n = 1, 2, 3$.

The problem with Fisher information

- ⋆ Recall that we are interested in bounding $H(F\|N)$, when

$$F = (f_1(G), ..., f_d(G)),$$

and the f_i's are non-linear transforms of the Gaussian field G. The problem is that, in such a general framework, there is actually **no available technique for properly bounding the relative Fisher information** $I(F\|N)$ without further heavy constraints on F.

- ⋆ This means that one has to be much more careful in studying the term

$$\int_0^x I(F_t\|N)\mathrm{d}t,$$

by suitably bounding $I(F_t\|N)$ in terms of the total variation distance. This can be done by exploiting an important inequality (due to **Carbery and Wright**, 2001), yielding bounds on the **small ball probabilities** of polynomial random variables. See Nourdin and Poly (2012).

THE PROBLEM WITH FISHER INFORMATION

- ⋆ Recall that we are interested in bounding $H(F\|N)$, when

$$F = (f_1(G), ..., f_d(G)),$$

and the f_i's are non-linear transforms of the Gaussian field G. The problem is that, in such a general framework, there is actually **no available technique for properly bounding the relative Fisher information** $I(F\|N)$ without further heavy constraints on F.

- ⋆ This means that one has to be much more careful in studying the term

$$\int_0^x I(F_t\|N)\mathrm{d}t,$$

by suitably bounding $I(F_t\|N)$ in terms of the total variation distance. This can be done by exploiting an important inequality (due to **Carbery and Wright**, 2001), yielding bounds on the **small ball probabilities** of polynomial random variables. See Nourdin and Poly (2012).

The problem with Fisher information

- ⋆ Recall that we are interested in bounding $H(F\|N)$, when

$$F = (f_1(G), ..., f_d(G)),$$

and the f_i's are non-linear transforms of the Gaussian field G. The problem is that, in such a general framework, there is actually **no available technique for properly bounding the relative Fisher information** $I(F\|N)$ without further heavy constraints on F.

- ⋆ This means that one has to be much more careful in studying the term

$$\int_0^x I(F_t\|N)\mathrm{d}t,$$

by suitably bounding $I(F_t\|N)$ in terms of the total variation distance. This can be done by exploiting an important inequality (due to **Carbery and Wright**, 2001), yielding bounds on the **small ball probabilities** of polynomial random variables. See Nourdin and Poly (2012).

The problem with Fisher information

- ⋆ Recall that we are interested in bounding $H(F\|N)$, when

$$F = (f_1(G), ..., f_d(G)),$$

and the f_i's are non-linear transforms of the Gaussian field G. The problem is that, in such a general framework, there is actually **no available technique for properly bounding the relative Fisher information** $I(F\|N)$ without further heavy constraints on F.

- ⋆ This means that one has to be much more careful in studying the term

$$\int_0^x I(F_t\|N)\mathrm{d}t,$$

by suitably bounding $I(F_t\|N)$ in terms of the total variation distance. This can be done by exploiting an important inequality (due to **Carbery and Wright**, 2001), yielding bounds on the **small ball probabilities** of polynomial random variables. See Nourdin and Poly (2012).

Theorem (Entropic 4^{th} moment theorem – Nourdin, Peccati and Swan (2014))

Let

$$F_n = (F_{1,n}, ..., F_{d,n}), \quad n \geq 1,$$

be a chaotic sequence such that F_n converges in distribution to $N = (N_1, ..., N_d) \sim \mathscr{N}(0, C)$, where $C > 0$. Set

$$\Delta_n := \mathbf{E}\|F_n\|^4_{\mathbb{R}^d} - \mathbf{E}\|N\|^4_{\mathbb{R}^d} > 0, \quad n \geq 1.$$

Then, as $n \to \infty$,

$$\boxed{H(F_n \| N) = O(1)\, \Delta_n \log \Delta_n.}$$

Theorem (Entropic 4^{th} moment theorem – Nourdin, Peccati and Swan (2014))

Let

$$F_n = (F_{1,n}, ..., F_{d,n}), \quad n \geq 1,$$

be a chaotic sequence such that F_n converges in distribution to $N = (N_1, ..., N_d) \sim \mathscr{N}(0, C)$, where $C > 0$. Set

$$\Delta_n := \mathbf{E}\|F_n\|^4_{\mathbb{R}^d} - \mathbf{E}\|N\|^4_{\mathbb{R}^d} > 0, \quad n \geq 1.$$

Then, as $n \to \infty$,

$$\boxed{H(F_n\|N) = O(1)\,\Delta_n \log \Delta_n.}$$

Theorem (Entropic 4^{th} moment theorem – Nourdin, Peccati and Swan (2014))

Let

$$F_n = (F_{1,n}, ..., F_{d,n}), \quad n \geq 1,$$

be a chaotic sequence such that F_n converges in distribution to $N = (N_1, ..., N_d) \sim \mathscr{N}(0, C)$, where $C > 0$. Set

$$\Delta_n := \mathbf{E}\|F_n\|_{\mathbb{R}^d}^4 - \mathbf{E}\|N\|_{\mathbb{R}^d}^4 > 0, \quad n \geq 1.$$

Then, as $n \to \infty$,

$$\boxed{H(F_n\|N) = O(1)\,\Delta_n \log \Delta_n.}$$

REMARKS

- ⋆ Of course, by Pinsker inequality, the above statement translates into a bound on the total variation distance.

- ⋆ In Ledoux, Nourdin and Peccati (2015): extensions to random vectors living in the eigenspaces of the generator $\mathcal{L}$ of a Markov semigroup $\{P_t\}$ – whenever some form of the Carbery-Wright inequality is available (some log-concave case, for instance).

Remarks

- Of course, by Pinsker inequality, the above statement translates into a bound on the total variation distance.

- In Ledoux, Nourdin and Peccati (2015): extensions to random vectors living in the eigenspaces of the generator $\mathcal{L}$ of a Markov semigroup $\{P_t\}$ – whenever some form of the Carbery-Wright inequality is available (some log-concave case, for instance).

An explicit example

Consider the sequence

$$V_n := \frac{1}{\sqrt{n}} \left(\sum_{k=1}^{n} (X_k^2 - 1), \sum_{k=1}^{n} (X_k^3 - 3X_k) \right), \quad n \geq 1,$$

where the Gaussian sequence X_k has autocorrelation $\varrho(k) \sim |k|^\alpha$, for some $\alpha < -2/3$ (this corresponds to the increments of a fractional Brownian motion of Hurst index $H < 2/3$). Let $N \sim \mathcal{N}(0, \mathbb{I}_2)$.

Then, for a suitable constant $\sigma > 0$,

$$H(\sigma^{-1} V_n \| N), \ \mathbf{TV}(\sigma^{-1} V_n, N)^2 = O(1) \frac{\log n}{n}.$$

AN EXPLICIT EXAMPLE

Consider the sequence

$$V_n := \frac{1}{\sqrt{n}} \left(\sum_{k=1}^{n} (X_k^2 - 1), \sum_{k=1}^{n} (X_k^3 - 3X_k) \right), \quad n \geq 1,$$

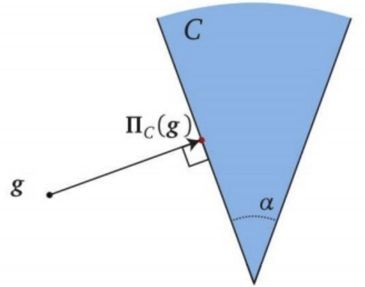
where the Gaussian sequence X_k has autocorrelation $\varrho(k) \sim |k|^\alpha$, for some $\alpha < -2/3$ (this corresponds to the increments of a fractional Brownian motion of Hurst index $H < 2/3$). Let $N \sim \mathcal{N}(0, \mathbb{I}_2)$.

Then, for a suitable constant $\sigma > 0$,

$$\boxed{H(\sigma^{-1} V_n \| N), \ \mathbf{TV}(\sigma^{-1} V_n, N)^2 = O(1) \tfrac{\log n}{n}.}$$

ANOTHER EXAMPLE

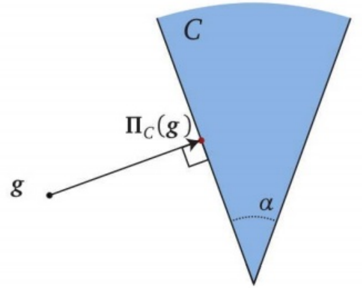
- ⋆ Closed convex cone $C \subset \mathbb{R}^d$; g = d-dimensional Gaussian vector; $\Pi_C(g)$ = metric projection of g onto C.



- ⋆ **Question** (Amelunxen, Lotz, McCoy and Tropp, 2013; Compressed Sensing) : *how far is the distribution of* $\|\Pi_C(g)\|^2$ *from that of a Gaussian random variable N?*

ANOTHER EXAMPLE

- ⋆ Closed convex cone $C \subset \mathbb{R}^d$; $g = d$-dimensional Gaussian vector; $\Pi_C(g)$ = metric projection of g onto C.



- ⋆ **Question** (Amelunxen, Lotz, McCoy and Tropp, 2013; Compressed Sensing) : *how far is the distribution of* $\|\Pi_C(g)\|^2$ *from that of a Gaussian random variable* N*?*

Another Example

⋆ **The HSI inequality** yields (under non-degeneracy): for large d,

$$\boxed{\mathbf{Ent}\,(N) - \mathbf{Ent}\big(\|\Pi_C(g)\|^2\big) \lesssim \frac{\log\delta}{\delta}}$$

where $\delta = \mathbb{E}\left(\|\Pi_C(g)\|^2\right)$ = 'statistical dimension' of the cone.

Another Example

⋆ **The HSI inequality** yields (under non-degeneracy): for large d,

$$\boxed{\mathbf{Ent}\,(N) - \mathbf{Ent}\big(\|\Pi_C(g)\|^2\big) \lesssim \frac{\log \delta}{\delta}}$$

where $\delta = \mathbb{E}\left(\|\Pi_C(g)\|^2\right)$ = ‘statistical dimension’ of the cone.

THANK YOU FOR YOUR ATTENTION !