Rapport de stage de M1

Martin Rakovsky, encadré par Max Fathi

Table des matières

Première partie
Présentation générale du déroulement du stage

Mon stage s'est déroulé du 8 février au 26 juin de l'année 2021, sous la direction de Max Fathi, au Laboratoire de Probabilités, Statistique et Modélisation. Je le remercie vivement pour son accueil, sa disponibilité et pour m'avoir montré un sujet de recherche passionnant. Je remercie également monsieur Giacomin pour m'avoir dirigé vers lui, et monsieur Mahfouf pour ses précieux conseils.

Désireux d'effectuer un stage qui mêlerait probabilités et analyse et devant l'incertitude, au mois de décembre, sur la possibilité de voyager en 2021, je me suis adressé à monsieur Giacomin pour demander un stage à Paris. Il m'a alors indiqué les travaux de Max Fathi. En lisant quelques articles, je suis alors tout de suite fasciné par le mélange entre probabilités et analyse dans ses travaux.
Après une première rencontre par écrans interposés, Max Fathi me propose plusieurs domaines d'études, avec un partage plus ou moins équilibré entre travail de recherche et travail bibliographique. C'est ainsi que je choisis comme objet d'étude les noyaux de Stein. Nous précisons ensuite ensemble les modalités du stage. J'apprends que je ne pourrai pas disposer d'un bureau au laboratoire et que la majeure partie du stage s'effectuera à distance, mais qu'il me sera possible de me rendre au laboratoire une fois par semaine pour échanger avec lui. Ces échanges devant un tableau s'avèreront très fructueux.
Dans un premier temps, mon travail alterne entre découverte de différents travaux autour de noyaux de Stein en lisant des articles et recherche personnelle sur un problème posé par mon maître de stage. Pendant nos échanges, je lui raconte ce que j'ai compris des différents articles, je lui montre où j'en suis dans ma recherche et éventuellement où je bloque. Sans pour autant faire le problème à ma place, il m'indique alors plusieurs ressources qui pourraient m'être utiles pour avancer.

Mes recherches ayant plutôt bien avancé après deux mois, Max Fathi me propose un petit intermède pour me montrer d'autres thèmes de recherche, avec plus ou moins de liens avec les noyaux de Stein, afin d'élargir mes horizons. Puis, j'ai à nouveau le choix entre un thème très orienté vers les probabilités, ou de continuer dans les noyaux de Stein, mais en restant alors dans un thème majoritairement d'analyse. Ayant pris goût à l'analyse pendant la première moitié du stage, je choisis la deuxième option, avec un problème qui m'occupera jusqu'à la fin du stage (et que je n'ai d'ailleurs pas pu achever).

Pendant le stage, Max Fathi me demande également d'assister au séminaire de probabilités du laboratoire. Celui-ci avait lieux intégralement en visio-conférence et était assez difficile à suivre pour moi. Les thèmes abordés étaient très variés et semblaient très intéressants, mais je n'ai pu en comprendre qu'une petite partie. Cette expérience était toutefois très formatrice et instructive.

En dehors du stage, j'ai eu la possibilité de suivre le cours d'intégrale stochastique à l'ENS et un cours d'analyse spectrale à Jussieu recommandé par mon maître de stage.

Deuxième partie

Brève description du travail mathématique

Le but du stage était de m'initier à certaines méthodes dites de Stein. Il s'agit d'un ensemble de méthodes permettant de comparer des mesures de probabilités. La méthode sur laquelle je me suis concentré est la méthode des noyaux de Stein.
Etant donnée une mesure μ sur $\mathbb{R}^d$ et un espace E de fonctions tests, un noyau de Stein pour μ est une fonction τ à valeurs matricielles vérifiant pour toute fonction test la formule d'intégration par partie suivante :

$$\int x \cdot \nabla f(x) d\mu = \int \langle \tau(x), \text{Hess } f(x) \rangle d\mu$$

L'idée est de se dire que deux mesures sont proches si elles vérifient presque la même formule d'intégration par partie, c'est-à-dire si leurs noyaux de Stein sont proches. Les domaines d'application sont divers, mais pendant le stage nous avons surtout utilisé cette notion pour comparer les mesures étudiées à la gaussienne standard sur $\mathbb{R}^d$, pour laquelle un noyau de Stein est l'identité, et d'en déduire des théorèmes Centraux Limite pour la distance de Wasserstein.

Il s'agit alors dans un premier temps de m'imprégner de cette notion à travers des exemples déjà existants dans la littérature. En amont, il est également nécessaire que je m'initie à certaines notions, comme les mesures vérifiant une inégalité de Poincaré ou de Sobolev-logarithmique ou encore la notion de semi-groupe de Markov et de générateur infinitésimal, qui sont des outils fréquemment rencontrés dans les articles.

Après cette initiation, je me vois proposé de déterminer un noyau de Stein pour les mesures μ à symétrie radiale et d'en déduire la vitesse de convergence de la somme normalisée de n variables aléatoires indépendantes de loi μ vers la gaussienne pour la distance de Wasserstein. En amont, mon maître de stage me fait lire d'autres articles sur les noyaux de Stein ou sur des méthodes similaires.

Pendant, la deuxième moitié du stage, on généralise la notion de noyaux de Stein en étudiant des noyaux d'ordre supérieur, c'est-à-dire des noyaux qui vérifient des formules d'intégration par parties plus évoluées. On compare alors la vitesse de convergence obtenue pour un noyau de Stein classique et pour un noyau d'ordre supérieur. Cette étude demeure inachevée par manque de temps.

Troisième partie
Contenu du stage

Dans toute la suite, on adoptera les conventions suivantes :

- Les mesures étudiées sont des mesures sur $\mathbb{R}^d$, avec $d \geqslant 1$ un entier.
- On note γ le mesure de la gaussienne standard sur $\mathbb{R}^d$.
- On note $x \cdot y$ le produit scalaire de deux éléments x et y de $\mathbb{R}^d$ et $|\cdot|$ la norme associée.
- On note $\langle A, B\rangle := \mathrm{Tr}(A^t B)$ le produit scalaire de Hilbert-Schmidt de deux matrices A et B de $\mathcal{M}_d(\mathbb{R})$ et $\|\cdot\|$ la norme associée.
- $\partial_i f$ désigne la dérivée partielle de f par rapport à la $i-$ème coordonnée (ou par rapport à la variable mise en indice).

1 Initiation

Il s'agit ici de présenter le contexte de notre étude. Avant de présenter la notion centrale de noyau de Stein d'une mesure et la motivation de son étude, on définit les notions d'inégalité de Poincaré et de semi-groupe d'opérateur. On se concentrera sur l'exemple canoniuqe de l'inégalité de Poincaré pour la gaussienne standard et le semi-groupe d'Ornstein-Uhlenbeck. Même si ces notions n'apparaissent pas explicitement dans les parties qui suivent, elles donnent souvent le contexte de l'étude et la motivation de nombreux résultats rencontrés dans les articles.

1.1 Premiers pas : l'inégalité de Poincaré et opérateurs

Les notions présentées ici sont détaillées dans [4].

1.1.1 Autour des inégalités de Poincaré

Définition 1.1. *Une mesure de probabilité sur $\mathbb{R}^d$ vérifie une inégalité de Poincaré sur un certain espace de fonctions tests s'il existe une constante C_p telle que pour toutes fonctions tests, on a l'inégalité suivante :*

$$\int f^2 d\mu \leqslant C_p \int |\nabla f|^2 d\mu$$

Remarque :

- Il est nécessaire que les moments d'ordre 1 et 2 de μ soient bien définis.
- L'espace des fonctions tests utilisé est souvent l'espace $H^1(\mu)$. Notons également qu'on peut se restreindre aux fonctions centrées pour μ, que l'on note $H^1_0(\mu)$.
- Le membre de gauche est bien sûr identifié comme la variance de f. Le membre de droite est parfois appelé *énergie* de la fonction f.

Une propriété importante de l'inégalité de Poincaré est qu'elle est stable par tensorisation.
Pour le voir, on commence par le lemme suivant :

Lemme 1.1. *Soit $(\Omega_i, \mathcal{F}_i, \mu_i)_{1\leqslant i\leqslant n}$ une famille d'espaces de probabilités et $(\Omega, \mathcal{F}, \mu^n)$ l'espace produit. Alors on a*

$$\mathrm{Var}_{\mu^n}(f) \leqslant \sum_{i=1}^{n} E_{\mu^n}(\mathrm{Var}_{\mu_i}(f))$$

Corollaire 1.1. *Soient $\mu_1, \dots, \mu_d$ des mesures de probabilités sur $\mathbb{R}^d$ vérifiant des inégalités de Poincaré de constantes respectives C_{p_i}. Soit μ la mesure produit des μ_i sur $\mathbb{R}^d$. Alors μ vérifie une inégalité de Poincaré de constante $C_p \leqslant \max\{C_{p_i}, 1 \leqslant i \leqslant d\}$.*

Preuve du corollaire. Cela découle du calcul suivant :

$$\begin{aligned} \mathrm{Var}_\mu(f) &\leqslant \textstyle\sum_{i=1}^d E_\mu(\mathrm{Var}_{\mu_i}(f)) \\ &\leqslant \textstyle\sum_{i=1}^d E_{\mu^n}(C_{p_i} E_{\mu_i}(|\partial_i f|^2)) \\ &= \left(\max_{1\leqslant i\leqslant d} C_{p_i}\right) E_\mu(|\nabla f|^2) \end{aligned}$$

□

Preuve du lemme. On procède par récurrence sur n. L'inégalité est triviale pour $n = 1$ et il suffit alors de la démontrer pour $n = 2$ pour obtenir l'hérédité. Mais pour $n = 2$, en posant $\mu := \mu^2$, l'inégalité se réécrit :

$$E_\mu(f - E_{\mu_1}(f))^2 \geqslant E_\mu(E_{\mu_2}(f) - E_\mu(f))^2$$

qui correspond à l'inégalité de Jensen pour la mesure μ_2. □

Cette tensorisation montre une certaine stabilité des inégalités de Poincaré pour les mesures produits, dont les constantes de Poincaré ne dépendent pas de la dimension.

Une mesure qui nous intéresse particulièrement dans le reste de l'étude est la mesure gaussienne, qui vérifie une inégalité de poincaré avec constante 1 :

Théorème 1.1 (Inégalité de Poincaré pour la Gaussienne)**.** *La mesure gaussienne standard γ sur $\mathbb{R}$ vérifie une inégalité de Poincaré de constante 1 pour la classe de fonctions $H_0^1(\gamma)$.*

La constant est optimale, elle est réalisée pour la fonction identité.
La preuve, détaillée dans [[4]], repose sur les idées suivantes :

1. Se ramener au cas $d = 1$ par le corollaire précédent.
2. Se restreindre à la famille des fonctions de classe C^2 à support compact, qui sont denses dans $H_0^1(\gamma)$.
3. Montrer une inégalité de Poincaré pour la mesure $\frac{1}{2}(\delta_{-1} + \delta_1)$
4. Conclure en utilisant le théorème Centrale-Limite et le lemme de tensorisation.

1.1.2 Semi-groupe de Markov et générateur infinitésimal

On présente la notion de semi-groupe de Markov, et plus précisément le sous-groupe d'Ornstein-Uhlenbeck ainsi que son générateur infinitésimal.

Définition 1.2. *Soit $\mathcal{A}$ un espace de fonctions tests muni d'une norme $|\cdot|$. Un semi-groupe de Markov est une famille $(P_t)_{t\geqslant 0}$ d'opérateurs linéaires bornés vérifiant :*

- $P_0 = I_d$
- *Pour toute fonction test f, l'application $t \mapsto P_t f$ est continue sur $[0, +\infty[$.*
- *Pour tout $t, s \geqslant 0$, $P_t \circ P_s = P_{t+s}$.*
- $P_t 1 = 1$ *pour tout t et $P_t f \geqslant 0$ si $f \geqslant 0$ pour tout t et toute fonction test f.*
- P_t *est une contraction pour tout t : pour toute fonction test, $|P_t f| \leqslant |f|$.*

Définition 1.3. *On appelle alors générateur infinitésimal du semi-groupe $(P_t)_{t\geqslant 0}$ l'opérateur L définit par la limite suivante, si elle existe :*

$$Lf := \lim_{t\to 0} \frac{1}{t}(P_t f - f)$$

L'opérateur L permet de considérer un semi-groupe sous représentation exponentielle $P_t = e^{tL}$. Il vérifie également la propriété fondamentale suivante :

Proposition 1.1. *Pour toute fonction test telle que Lf est bien définie :*

$$\partial_t P_t f = P_t L f = L P_t f$$

Cette propriété permet de reconstruire le semi-groupe de Markov à partir de son générateur infinitésimal. En effet, étant une fonction test f, on peut retrouver $P_t f$ comme la solution de l'équation aux dérivées partielles avec condition initiale en temps suivante :

$$\begin{cases} \partial_t f(t,x) &= Lf(t,x) \quad \forall (t,x) \in \mathbb{R}_+ \times \mathbb{R}^d \\ f(0,x) &= f(x) \quad\quad\;\; \forall x \in \mathbb{R}^d \end{cases}$$

Le semi-groupe qui va particulièrement nous intéresser ici est le semi groupe suivant :

Définition 1.4. *Soit Y une variable aléatoire gaussienne standard. Le semi-groupe d'Ornstein-Uhlenbeck est le semi-groupe d'opérateur (P_t) défini par*

$$P_t f(x) := E[f(e^{-t}x + \sqrt{1-e^{-2t}}Y)]$$

Plusieurs propriétés nous permettent d'effleurer l'intérêt d'un tel semi-groupe. Tout d'abord, $P_0 f = f$ et $\lim_{t\to+\infty} P_t f = E_\gamma[f]$. On peut donc passer de façon différentiable de f à sa valeur moyenne. Cette propriété est très importante pour nous dans la mesure où de nombreuses preuves rencontrées dans les articles utilisent le sous-groupe d'Ornstein-Uhlenbeck pour réaliser une interpolation entre une variable aléatoire et une gaussienne standard.

Le générateur infinitésimal du semi-groupe d'Ornstein-Uhlenbeck est l'opérateur suivant :

$$Lf(x) = \Delta f(x) - x \cdot \nabla f(x)$$

Notons d'emblée que $E_\gamma[Lf(X)] = 0$ par intégration par partie. Il s'agit d'un cas particulier d'une propriété plus générale d'invariance : pour toute fonction f, $E_\gamma[P_t f(X)] = E_\gamma[f(X)]$.

1.2 Présentation des noyaux de Stein

Définition 1.5. *On se donne* μ *une mesure de probabilité sur* $\mathbb{R}^d$ *et un ensemble* E *de fonction tests de* $\mathbb{R}^d$ *dans* $\mathbb{R}^d$*. On appelle noyau de Stein de la mesure* μ *toute fonction* $\tau_\mu : \mathbb{R}^d \mapsto \mathcal{M}_d(\mathbb{R})$ *vérifiant l'égalité suivante pour toute fonction test* f *de* E *:*

$$\int x \cdot f(x) d\mu = \int \langle \tau_\mu(x), \nabla f(x) \rangle d\mu$$

Remarque :

- L'espace des fonctions test est souvent $H_0^1(\mathbb{R})$, qui convient dès que μ admet un moment d'ordre 2. Dans la suite, on supposera implicitement sauf mention du contraire que l'on est dans ce cadre.
- Une autre façon, parfois plus commode, mais équivalente à la première est de prendre comme fonctions test des fonctions de $\mathbb{R}^d$ dans $\mathbb{R}$ et de définir les noyaux de Stein comme vérifiant

$$\int x \cdot \nabla f(x) d\mu = \int \langle \tau_\mu(x), \text{Hess } f(x) \rangle d\mu$$

- On peut voir le noyau Stein comme une fonction permettant d'avoir une formule d'intégration par partie pour la mesure μ.

1.3 Motivations de l'étude

La principale motivation de l'introduction des noyaux de Stein est le lemme suivant, de Stein :

Lemme 1.2. *La mesure gaussienne est la seule mesure admettant l'identité comme noyau de Stein. Autrement dit, la mesure gaussienne standard est la seule mesure de probabilité vérifiant pour toute fonction* f *de* $H^1(\mathbb{R})$ *:*

$$\int x \cdot f d\mu = \int \mathrm{div}(f) d\mu$$

On introduit donc la quantité suivante, appelée *discrépance* de Stein d'une mesure μ :

$$S(\mu)^2 := \inf \left\{ \int ||I_d - \tau_\mu||^2 d\mu | \tau_\mu \text{ est un noyau de Stein pour } \mu \right\}$$

qui estime si le noyau de Stein trouvé est proche du noyau de la gaussienne. De cette estimation, on espère déterminer si la mesure μ est elle-même proche de la gaussienne.
Cette étude de la proximité entre deux mesures se fait en deux étapes :

1. Le recours à la distance de Wasserstein, qui intervient principalement dans la théorie du transport optimal, pour comparer deux mesures. Celle-ci se définit comme suit :

   Définition 1.6. *Soit* $p \geqslant 1$*. Soient* μ *et* ν *deux mesures de probabilités admettant chacune un moment d'ordre* p*. La distance* $p-$*ème de Wasserstein des mesures* μ *et* ν *est la quantité*

$$W_p(\mu, \nu) := \left(\inf_{\pi \in \Gamma(\mu,\nu)} \int |x-y|^p d\pi(x,y) \right)^{1/p}$$

   où $\Gamma(\mu, \nu)$ *est l'ensemble des mesures* π *dont la projection sur la première coordonnée est* μ *et la projection sur la seconde coordonnée est* ν*.*

En pratique, on ne considérera dans le suite que la distance $W_2(\mu, \nu)$. La convergence d'une famille de deux mesures pour la distance de Wasserstein implique la convergence de faible de la famille. Ainsi, les résultats de convergence que l'on obtiendra par la suite sont légèrement plus forts que le théorème Centrale Limite classique.

2. L'utilisation de la discrépance de Stein, dont l'une des propriétés principales est l'inégalité suivante :

$$W_2(\mu, \gamma)^2 \leqslant S(\mu)^2$$

dont la preuve utilise une interpolation entre une variable de loi μ et une gaussienne standard via le semi-groupe d'Ornstein-Uhlenbeck. On peut en trouver un exemple dans [5].
Ainsi, dans la suite, il nous suffira d'établir des majorations sur la discrépance de Stein pour obtenir des vitesses de convergence pour la distance de Wasserstein.

2 Quelques exemples de noyaux de Stein pour diverses mesures

La section suivante est un corpus mettant en relation quelques articles lus pendant le stage, notamment les articles [2], [3], [5]. L'exemple le plus instructif est celui des mesures vérifiant une inégalité de Poincaré. C'est l'occasion de présenter le protocole à appliquer pour, à partir d'un noyau de Stein pour une mesure μ, déterminer la vitesse de convergence de μ^n vers la gaussienne standard, où μ^n est la loi de la somme normalisée de n variables indépendantes de loi μ. On poursuit ensuite avec l'exemple de mesures possédant un noyau de Stein positif.

2.1 Cas des mesures vérifiant une inégalité de Poincaré

Dans cette partie, on présente quelques conditions d'existence d'un noyau de Stein pour une mesure μ.
Tout d'abord, en prenant $f = 1$ comme fonction test, on constate qu'il est nécessaire d'être centrée pour une mesure μ admettant un noyau de Stein. Cela n'est en revanche pas une condition suffisante, et il n'est pas difficile de se convaincre que la mesure $\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$ n'admet pas de noyau de Stein.
Dans la suite de cette sous-section, on suppose que μ est absolument continue par rapport à la mesure de Lebesgues et que μ est centrée et admet un moment d'ordre 2, c'est-à-dire que $\int |x|^2 d\mu < \infty$.

Théorème 2.1. *Si μ vérifie une inégalité de Poincaré de constante C_p, alors il existe une fonction g de $H^1(\mu)$ d'espérance nulle telle que ∇g est un noyau de Stein pour la mesure μ.*

De plus, on a

$$\int ||\nabla g||^2 d\mu \leqslant C_p \int |x|^2 d\mu$$

si bien que

$$S(\mu)^2 \leqslant (C_p - 2) \int |x|^2 d\mu + d$$

Démonstration. L'existence de l'inégalité de Poincaré pour μ rend propice l'utilisation du théorème de Lax-Milgram. Plus précisément, en notant $H_0^1(\mu)$ l'ensemble des fonctions de $H^1(\mu)$ d'espérance nulle, la forme bilinéaire

$$(f_1, f_2) \in H_0^1(\mu)^2 \longmapsto \int \langle \nabla f_1, \nabla f_2 \rangle d\mu$$

est coercive de constante $1/C_p$. D'autre part, la forme linéaire

$$f \in H_0^1(\mu) \longmapsto \int x \cdot f d\mu$$

est continue (d'après Cauchy-Schwarz et puisque μ admet un moment d'ordre 2 fini). On dispose donc d'une fonction g de $H^1(\mu)$ telle que pour toute fonction f de $H_0^1(\mu)$:

$$\int \langle \nabla g, \nabla f \rangle d\mu = \int x \cdot f d\mu$$

et ∇g est un noyau de Stein pour μ.

Puisque la forme bilinéaire est symétrique, le théorème de Lax-Milgram donne également que g minimise la quantité $\frac{1}{2} \int ||\nabla f||^2 d\mu - \int x \cdot f d\mu$. On déduit que

$$\begin{aligned}
-\frac{1}{2} \int ||\nabla g||^2 d\mu &= \frac{1}{2} \int ||\nabla g||^2 d\mu - \int x \cdot g d\mu \\
&\geqslant \frac{1}{2C_p} \int |g|^2 d\mu - \left(\int |x|^2 d\mu \right)^{1/2} \left(\int |g|^2 d\mu \right)^{1/2} \\
&\geqslant -\frac{C_p}{2} \int |x|^2 d\mu
\end{aligned}$$

On obtient donc la première inégalité du théorème.
Cette inégalité nous fournit un contrôle pour la discrépance de Stein :

$$\begin{aligned}
S(\mu)^2 &\leqslant \int ||\nabla g - I_d||^2 d\mu \\
&= \int ||\nabla g||^2 d\mu + \int ||I_d||^2 d\mu - 2 \int \rangle \nabla g, I_d \langle d\mu \\
&\leqslant C_p \int |x|^2 d\mu + d \int d\mu - 2 \int x \cdot x d\mu \\
&= (C_p - 2) \int |x|^2 d\mu + d
\end{aligned}$$

ce qui complète le théorème. □

2.2 Utilisation pour évaluer la distance de Wasserstein

Soit X une variable aléatoire dont la loi μ est centrée réduite et satisfait une inégalité de Poincaré et soit μ_n la loi de $\sum X_i/\sqrt{n}$, où les X_i sont indépendantes de loi μ.

Dans cette partie, on cherche à obtenir la vitesse de convergence de la quantité $W_2(\mu_n, \gamma)$. Comme annoncé, on va obtenir cette convergence grâce à la convergence de la discrépance de Stein $S(\mu_n)$. On se sert pour cela du lemme suivant :

Lemme 2.1. *Si μ admet un noyau de Stein, alors pour tout n, μ_n admet un noyau de Stein et on a l'inégalité :*

$$S(\mu_n)^2 \leqslant \frac{1}{n} S(\mu)^2$$

Démonstration. On note $S_n = \sum X_i/\sqrt{n}$. La clé est de remarquer que si τ est un noyau de Stein pour la loi de μ, alors la fonction

$$\tau_n(s_n) := E[\tau(X_1)|S_n = s_n]$$

est un noyau de Stein pour μ_n. En effet, soit f une fonction test à valeurs dans $\mathbb{R}^d$. On a alors la relation

$$\nabla_{x_1} f = \frac{1}{\sqrt{n}} \nabla_{s_n} f$$

si bien que

$$\begin{aligned}
E[S_n \cdot f(S_n)] &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} E[X_i \cdot f(S_n)] \\
&= \sqrt{n} E[X_1 \cdot f(S_n)] \\
&= \sqrt{n} E\left[E[X_1 \cdot f(S_n)|S - X_1/\sqrt{n}]\right] \\
&= \sqrt{n} E\left[E[\langle \tau(X_1), \nabla_{X_1} f(S_n)\rangle | S - X_1/\sqrt{n}]\right] \\
&= E\left[\langle \tau(X_1), \nabla_{S_n} f(S_n)\rangle\right]
\end{aligned}$$

ce qui donne le résultat annoncé.
Pour en déduire l'inégalité, on constate que pour tout noyau de Stein de μ, on a par l'inégalité de Jensen :

$$S(\mu_n)^2 \leqslant E\left[\|\tau_{\mu_n}(S_n) - I_d\|^2\right] = E\left[\|E[\tau(X_1)|S_n] - I_d\|^2\right] \leqslant \frac{1}{n^2} E\left[E\left[\left\|\sum_{i=1}^{d} (\tau(X_k) - I_d)\right\|^2 |S_n\right]\right]$$

Puis, en utilisant que les variables $\tau(X_k) - I_d$ sont mutuellement indépendantes et centrées :

$$\frac{1}{n^2} E\left[\left\|\sum_{i=1}^{d} (\tau(X_k) - I_d)\right\|^2\right] = \frac{1}{n^2} \sum_{i=1}^{d} E\left[\|\tau(X_k) - I_d\|^2\right] = \frac{1}{n} E\left[\|\tau(X_1) - I_d\|^2\right]$$

On passe alors à l'infimum pour obtenir l'inégalité $S(\mu_n)^2 \leqslant \frac{1}{n} S(\mu)^2$. □

De ce lemme et de la borne obtenue plus haut sur $S(\mu)^2$ dans le cas où μ vérifie une inégalité de Poincaré, on déduit :

Corollaire 2.1. *Pour une mesure μ normalisée par $\int |x|^2 d\mu = d$ et vérifiant une inégalité de Poincaré de constante C_p, on a pour tout entier n l'inégalité*

$$W_2(\mu_n, \gamma)^2 \leqslant S(\mu_n)^2 \leqslant \frac{1}{n} S(\mu)^2 \leqslant \frac{d(C_p - 1)}{n}$$

Le contrôle en $O(1/\sqrt{n})$ de la distance de Wasserstein est connu pour être optimal.

2.3 Détour par les mesures log-concaves

Il est légitime de se demander dans quelles conditions on peut espérer une réciproque du théorème obtenu plus haut, c'est-à-dire si l'existence d'un noyau de Stein pour une mesure implique que celle-ci vérifie une inégalité de Poincaré.
Cette question est l'occasion d'étudier les noyaux de Stein d'une nouvelle famille de mesures. Notamment, il est grand temps de s'intéresser au cas où la mesure μ est une mesure sur $\mathbb{R}$.

Théorème 2.2. *Soit μ une mesure centrée sur $\mathbb{R}$ admettant une densité p à support connexe. Alors μ admet un unique noyau de Stein (modulo les ensembles de mesure nulle) donné par la formule suivante :*

$$\tau(x) = \frac{1}{p(x)} \int_x^\infty y p(y) dy$$

Un confort offert par la dimension 1 est que ce noyau est toujours positif, tandis qu'un noyau de Stein en dimension supérieure n'est pas forcément une matrice symétrique positive. En effet, on a bien $\tau(x) \geqslant 0$ si $x \geqslant 0$, puis, si $x \leqslant 0$, on remarque que

$$\int_0^\infty y p(y) dy = -\int_{-\infty}^x y p(y) dy \geqslant 0$$

Cette positivté va permettre de motiver l'extension du travail qui suit au cas des mesures sur $\mathbb{R}^d$ dont le noyau de Stein est défini positif.
En dimension 1, l'existence d'un noyau de Stein implique l'existence d'une inégalité de Poincaré, comme l'indique le théorème suivant :

Théorème 2.3. *Soit X une variable aléatoire de loi μ centrée admettant une densité p à support connexe. Soit τ l'unique noyau de Stein de la mesure μ. Alors la mesure μ vérifie une inégalité de Poincaré à poids, c'est-à-dire que pour tout fonction test f, on a*

$$Var[f(X)] \leqslant E[\tau(X)(f'(X))^2]$$

Démonstration. Dans la suite, on note $F(x)$ la fonction de répartition de p. On introduit également deux nouvelles notations :
On pose $k(x, y) = F(x \wedge y) - F(x)F(y)$ et pour une fonction test f, on pose $\mathcal{L}f(x) = \int_{-\infty}^x k(x, y) f'(y) dy$.
On utilise par la suite l'identité suivante, présentée dans [3], valable pour toutes fonctions tests f et g :

$$Cov(f, g) = \int f'(x) k(x, y) g'(y) dx dy$$

Cette égalité implique, combinée avec Cauchy-Schwarz :

$$Var[f(X)] = \int f'(x) \mathcal{L}f(x) dx \leqslant \sqrt{\int \tau(x) f'(x)^2 p(x) dx} \sqrt{\int \tau(x) \left(\frac{p(x)^{-1} \mathcal{L}f(x)}{\tau(x)} \right)^2 p(x) dx}$$

Puis, en remarquant que $\tau(x) = p(x)^{-1}\mathcal{L}\mathrm{Id}(x)$, on obtient que

$$\left(\frac{p(x)^{-1}\mathcal{L}f(x)}{\tau(x)}\right)^2 = \left(\int f'(y)\frac{k(x,y)}{\int k(z,x)dz}dy\right)^2 \leqslant \int f'(y)^2\frac{k(x,y)}{\int k(x,z)}dz = \frac{1}{\tau(x)p(x)}\int f'(y)^2 k(x,y)dy$$

où on a utilisé l'inégalité de Jensen pour la fonction carrée. Si bien que

$$\int \tau(x)\left(\frac{p(x)^{-1}\mathcal{L}f(x)}{\tau(x)}\right)^2 p(x)dx \leqslant \int f'(y)^2 \int k(x,y)dxdy = \int f'(y)^2\tau(y)p(y)dy$$

On obtient l'inégalité désirée en réécrivant les termes $\int f'(x)^2\tau(x)p(x)dx$ comme $E[\tau(X)f'(X)^2]$.

□

Dans la preuve qui précède, on a bien vu l'intérêt d'un noyau positif, qu'on peut alors passer à la racine.

Dans le cas de la dimension supérieure, on ne peut donc espérer obtenir un théorème similaire seulement si la mesure étudiée admet un noyau à valeurs dans les matrices symétriques définies positives. Ceci nous motive à nous intéresser aux noyaux qui sont la hessienne de fonctions convexes. C'est le cas des mesures log-concaves, c'est-à-dire des mesure admettant une densité de la forme $e^{-\varphi}$, où φ est une fonction convexe.
On dispose notamment du résultat suivant :

Théorème 2.4. *Soit μ une mesure de densité ρ. On suppose que l'équation aux dérivées partielles suivante :*

$$e^{-\varphi} = \rho(\nabla\varphi)\det(\mathrm{Hess}(\varphi))$$

admet une solution φ convexe de classe C^2 strictement positive sur $\mathbb{R}^d$. Si on note $\mathcal{F}^\star$ la transformée de Legendre de la fonction f, alors $\mathrm{Hess}(\varphi)(\nabla\varphi^\star)$ est un noyau de Stein pour la mesure μ. On a alors

$$S(\mu)^2 \leqslant \int ||\mathrm{Hess}(\varphi) - I_d||^2 e^{-\varphi}dx$$

Remarques :

1. Ce résultat est en fait motivé par une combinaison de la théorie du transport optimal et de l'étude des équations de Monge-Ampère. En effet, l'équation aux dérivées partielles sur φ se traduit par le fait que μ est l'image de la mesure à densité e^{φ} par la fonction $\nabla\varphi$. Ainsi, au lieu de fixer deux mesures et de chercher la fonction qui transporte l'une vers l'autre avec un coût minimal, on fixe ici seulement la cible et on cherche une source et un transport qui soient liés.
2. On dispose en fait de plusieurs résultats produisant des mesures vérifiant les hypothèses ci-dessus. Le premier est le suivant, que l'on retrouve dans [8] :

   Théorème 2.5. *(Cordero-Erausquin et Klartag) Soit μ une mesure centrée admettant un premier moment et dont le support n'est pas inclus dans un hyperplan. Alors il existe une fonction convexe φ telle que μ est l'image de la mesure de densité $e^{-\varphi}$ par $\nabla\phi$.*

   Ce théorème n'assure cependant pas de régularité sur φ. On le complète par le résultat suivant, établi dans [9] :

Théorème 2.6. *(Berman et Berndston) Si μ est à support compact, ouvert et convexe, si ρ est lisse et si il existe une constante $C > 0$ telle que $C \geqslant \rho \geqslant C^{-1}$, alors φ est lisse et strictement positive sur $\mathbb{R}^d$.*

En particulier, si μ vérifie les hypothèses du théorème précédent, alors μ admet le noyau de Stein annoncé.

3. Le transport inverse, de μ vers $e^{-\varphi}$, est bien connu en transport otpimal, et la seule fonction réalisant le transport optimal est en fait $\nabla\varphi^\star$.

Démonstration. Tout d'abord, une propriété utile pour la suite est que $\nabla\varphi^\star$ et $\nabla\varphi$ sont inverses l'une de l'autre.
Soit maintenant f une fonction test. En posant le changement de variable $y = \nabla\varphi^\star(x)$ et en utilisant l'équation satisfaite par ρ, on a

$$\int x \cdot f(x)\rho(x)dx = \int \nabla\varphi(y) \cdot f(\nabla\varphi(y))e^{-\varphi(y)}dy$$

Puis, en intégrant par parties, (en intégrant $\nabla\varphi e^{-\varphi}$ et en différenciant $f \circ \nabla\varphi$) :

$$\int \nabla\varphi(y) \cdot g(y)e^{-\varphi(y)}dy = \int \langle \nabla^2\varphi(y), \nabla f(\nabla\varphi(y))\rangle e^{-\varphi(y)}dy$$

On rééffectue un changement de variable pour retrouver

$$\int \langle \nabla^2\varphi(y), \nabla f(\nabla\varphi(y))\rangle e^{-\varphi(y)}dy = \int \langle \mathrm{Hess}(\varphi)\nabla\varphi^\star(x), \nabla f(x)\rangle \rho(x)dx$$

ce qui donne l'égalité voulue.

Le changement de variable permet également d'obtenir la borne sur la discrépance de Stein. □

Klartag a réussi à obtenir dans [10] une borne explicite pour le majorant de la discrépance, qui ne dépend que de μ et plus de φ :

Théorème 2.7. *Avec les notations précédentes, on a pour tout θ de $\mathcal{S}^{d-1}$:*

$$\int (\mathrm{Hess}(\varphi)(\nabla\varphi^\star)\theta \cdot \theta)^2 \, d\mu \leqslant 256 \left(\int (x \cdot \theta)^2 d\mu \right)^2$$

Une borne analogue existe si l'on s'intéresse à la distance $p-$ème de Wasserstein pour p général. Puisque cette borne ne dépend pas de φ, on déduit de même que pour le cas où la mesure vérifie une inégalité de Poincaré le résultat de convergence suivant :

$$W_2(\mu_n, \gamma) \leqslant C\frac{d}{\sqrt{n}}$$

où μ_n est la loi de somme normalisée de n variables indépendantes de loi μ.
Comme annoncé, on montre désormais qu'une telle mesure vérifie une inégalité de Poincaré. L'idée est toujours de transporter μ vers une mesure log-concave, puis d'utiliser que les mesures log-concaves vérifie l'inégalité de Poincaré suivante, pour toute fonction test f :

$$\int f^2 e^{-V}dx \leqslant \int \langle (\mathrm{Hess}\ V)^{-1}\nabla f, \nabla f\rangle e^{-V}dx$$

Ainsi, si f est une fonction test :

$$\begin{aligned}
\int |f|^2 d\mu &= \int |f(\nabla\varphi)|^2 e^{-\varphi} dx \\
&\leqslant \int \langle (\text{Hess } \varphi)^{-1} \nabla^2\varphi \nabla f(\nabla\varphi), \nabla^2\varphi\nabla f(\nabla\varphi)\rangle e^{-\varphi} dx \\
&= \int \langle \nabla f(\nabla\varphi), \text{Hess } \varphi \nabla f(\nabla\varphi)\rangle e^{-\varphi} dx \\
&= \int \langle \tau \nabla f, \nabla f\rangle d\mu
\end{aligned}$$

Il est à noter que $\tau = (\text{Hess } \varphi)^{-1}$, et donc on retrouve bien une inégalité de Poincaré similaire à celle dont on dispose pour les mesure log-concaves, ce qui permet de généraliser le cas de la dimension 1 et fournit une réciproque au résultat établi en début de section.

3 Noyau de Stein pour les mesures à symétrie radiale

L'objectif de cette section est d'établir un noyau de Stein pour les mesures μ sur $\mathbb{R}^d$ invariantes par rotation.
On commence par établir l'expression d'un noyau de Stein pour la mesure uniforme sur la sphère de rayon r. Puis, on déduit l'expression d'un noyau de Stein dans le cas général en exprimant les mesures radiales comme mixtures de mesures uniformes sur le cercle. Enfin, comme dans la section précédente, on utilise ce noyau de Stein pour majorer la distance de Wasserstein entre certaines mesures radiales et la gaussienne standard.

Dans toute la suite, on notera $\mathbb{S}_r^{d-1}$ la sphère centrée en 0 et de rayon r dans $\mathbb{R}^d$ et μ_r la mesure uniforme sur $\mathbb{S}_r^d$. Dans le cas de $\mathbb{R}^2$, on écrira plus simplement $\mathbb{S}_r$.

3.1 La mesure uniforme sur la sphère dans le cas de la dimension 2

On commence par regarder un cas simple et facile à visualiser, le cas de la sphère en dimension 2 (du cercle quoi). On produit ici un noyau de Stein pour μ_r.

Théorème 3.1. *Avec les notations introduites, la fonction*

$$\begin{array}{cccc}
\tau_{\mu_r} : & \mathbb{R}^2 & \longrightarrow & \mathcal{M}_2(\mathbb{R}) \\
 & (x,y) & \longmapsto & \begin{pmatrix} y^2 & -xy \\ -xy & x^2 \end{pmatrix}
\end{array}$$

est un noyau de Stein de la mesure μ_r vue comme une mesure sur $\mathbb{R}^2$.

Démonstration. Soit $c : t \mapsto (r\cos(t), r\sin(t))$ une paramétrisation du cercle. Soit φ une fonction de classe C^2 de $\mathbb{R}^2$ dans $\mathbb{R}^2$. Puisque c est fermée et φ lisse :

$$0 = \int_c \varphi'' d\mu_r = \int_0^{2\pi} (\varphi \circ c(t))'' dt$$

D'autre part

$$(\varphi\circ c)'(t) = c_x'(t)\partial_x\varphi\circ\gamma(t) + c_y'(t)\partial_y\varphi\circ c(t) = c'(t)\cdot\nabla(\varphi\circ c)(t)$$

et

$$
\begin{aligned}
(\varphi \circ c)''(t) &= c''(t) \circ \nabla(\varphi \circ c)(t) + \mathrm{Tr}\begin{pmatrix} c'_x(t)^2 & c'_x(t)c'_y(t) \\ c'_x(t)c'_y(t) & c'_y(t)^2 \end{pmatrix}\begin{pmatrix} \partial^2_{x,x}\varphi \circ c(t) & \partial^2_{x,y}\varphi \circ c(t) \\ \partial^2_{x,y}\varphi \circ c(t) & \partial^2_{y,y}\varphi \circ c(t) \end{pmatrix} \\
&= -c(t) \cdot \nabla(\varphi \circ c)(t) + \mathrm{Tr}\begin{pmatrix} c'_x(t)^2 & c'_x(t)c'_y(t) \\ c'_x(t)c'_y(t) & c'_y(t)^2 \end{pmatrix}\mathrm{Hess}(\varphi)(c(t))
\end{aligned}
$$

On déduit que

$$
0 = -\int_{\mathbb{S}_r} z \cdot \nabla\varphi(z)d\mu_r + \int_{\mathbb{S}_r} \mathrm{Tr}\left(\begin{pmatrix} y^2 & -xy \\ -xy & x^2 \end{pmatrix}^t \mathrm{Hess}(\varphi)(x,y)\right)d\mu_r
$$

ce qui achève la preuve. □

Constatons avec enthousiasme que l'expression de τ_r est bien invariante par rotation, ne dépend pas de r et semble suffisamment sympathique pour admettre une généralisation en dimension supérieure.

3.2 La mesure uniforme sur la sphère dans le cas général ($d \geqslant 3$)

La preuve s'adapte bien à la dimension supérieure, ce qui nous offre le résultat suivant.

Théorème 3.2. *Soit $\mathbb{S}_r^{d-1}$ la sphère de centre 0 de rayon r et de dimension $d-1$. Soit μ_r la mesure uniforme sur $\mathbb{S}_r^{d-1}$. Alors la fonction :*

$$
\begin{aligned}
\tau_{\mu_r} : \mathbb{R}^d &\longrightarrow \mathcal{M}_d(\mathbb{R}) \\
x &\longmapsto \frac{1}{d-1}\begin{pmatrix} x_2^2 + x_3^2 + \ldots + x_n^2 & -x_1x_2 & \cdots & -x_1x_n \\ -x_1x_2 & x_1^2 + x_3^2 + \ldots + x_n^2 & \cdots & -x_2x_n \\ \vdots & \vdots & \ddots & \vdots \\ -x_1x_n & -x_2x_n & \cdots & x_1^2 + x_2^2 + \ldots + x_{n-1}^2 \end{pmatrix}
\end{aligned}
$$

est un noyau de Stein pour μ_r.

Démonstration. On commence par établir un noyau pour la mesure uniforme sur S^{d-1} la sphère de rayon 1, en utilisant la même méthode que dans le cas $d = 2$. Pour cela, on utilise à nouveau que pour toute fonction φ lisse :

$$
0 = \int_{\mathbb{S}^{d-1}} \Delta_{\mathbb{S}^{d-1}}\varphi d\mu_1
$$

On doit ensuite calculer $\Delta_{\mathbb{S}^{d-1}}$. On adopte la méthode naïve : on prolonge notre fonction φ sur un voisingae de la sphère et on étudie $\hat{\varphi}(x) := \varphi\left(\frac{x}{||x||}\right)$, dont le laplacien correspond au laplacien sphérique de φ. On a alors :

$$
\partial_i\hat{\varphi}(x) = \frac{1}{||x||}\partial_i\varphi(x) - \frac{1}{||x||^3}\sum_{k=1}^{d} x_i x_k \partial_k\varphi(x)
$$

et

$$\partial^2_{i,i}\hat{\varphi}(x) = \frac{1}{||x||}\partial^2_{i,i}\varphi(x) - 3\frac{x_i}{||x||^3}\partial_i\varphi(x) - \sum_{k=1}^{d}\frac{x_i x_k}{||x||^3}\partial^2_{i,k}\varphi(x) - \sum_{k\neq i}^{d}\frac{x_k}{||x||^3}\partial_k\varphi(x) + 3\sum_{k=1}^{d}\frac{x_k x_i^2}{||x||^5}\partial_k\varphi(x)$$

On va sommer chaque terme sur i :

$$\begin{array}{rclcl}
\displaystyle\sum_{i=1}^{d}\frac{1}{||x||}\partial^2_{i,i}\varphi(x) & = & \displaystyle\frac{1}{||x||}\Delta_{\mathbb{R}^d}\varphi(x) & = & \displaystyle\frac{1}{||x||^3}\mathrm{Tr}\left(\mathrm{Hess}(\varphi)(x)^t||x||^2 I_d\right) \\
\displaystyle -3\sum_{i=1}^{d}\frac{x_i}{||x||^3}\partial_i\varphi(x) & = & & & \displaystyle -3\frac{1}{||x||^3}x\cdot\nabla\varphi(x) \\
\displaystyle -\sum_{i=1}^{d}\sum_{k=1}^{d}\frac{x_i x_k}{||x||^3}\partial^2_{i,k}\varphi(x) & = & & & -\mathrm{Tr}(\mathrm{Hess}(\varphi)(x)^t[x\otimes x]) \\
\displaystyle -\sum_{i=1}^{d}\sum_{k\neq i}^{d}\frac{x_k}{||x||^3}\partial_k\varphi(x) & = & & & \displaystyle -(d-1)\frac{1}{||x||^3}x\cdot\nabla\varphi(x) \\
\displaystyle 3\sum_{i=1}^{d}\sum_{k=1}^{d}\frac{x_k x_i^2}{||x||^5}\partial_k\varphi(x) & = & \displaystyle 3\sum_{k=1}^{d}\frac{x_k}{||x||^5}\partial_k\varphi(x)\underbrace{\sum_{i=1}^{d}x_i^2}_{=||x||^2} & = & \displaystyle 3\frac{1}{||x||^3}x\cdot\nabla\varphi(x)
\end{array}$$

On trouve donc :

$$\begin{aligned}
0 &= \int_{\mathbb{S}^{d-1}}\Delta_{\mathbb{S}^{d-1}}\varphi\, d\mu_1 \\
&= \int_{\mathbb{S}^{d-1}}\left[\frac{1}{||x||^3}\mathrm{Tr}\left(\mathrm{Hess}(\varphi)(x)^t||x||^2 I_d\right) - \mathrm{Tr}(\mathrm{Hess}(\varphi)(x)^t[x\otimes x])\right]d\mu_1(x) \\
&\quad -(d-1)\int_{\mathbb{S}^{d-1}}\frac{1}{||x||^3}x\cdot\nabla\varphi(x)\,d\mu_1(x) \\
&\quad +\underbrace{\int_{\mathbb{S}^{d-1}}3\frac{1}{||x||^3}x\cdot\nabla\varphi(x) - 3\frac{1}{||x||^3}x\cdot\nabla\varphi(x)\,d\mu_1(x)}_{=0} \\
&= (d-1)\int_{\mathbb{S}^{d-1}}\mathrm{Tr}(\mathrm{Hess}(\varphi(x))^t\tau_{\mu_1}(x))\,d\mu_1(x) - (d-1)\int_{\mathbb{S}^{d-1}}\frac{1}{||x||^3}x\cdot\nabla\varphi(x)\,d\mu_1(x)
\end{aligned}$$

ce qui donne bien le résultat voulu pour $\mathbb{S}^{d-1}$.
Les coefficients de la matrice τ_{μ_1} sont homogènes de degré 2, on va donc facilement obtenir une invariance d'échelle :

$$\begin{aligned}
\int_{\mathbb{S}_r^{d-1}} x \cdot \nabla\varphi(x) d\mu_r(x) &= \int_{\mathbb{S}^{d-1}} (rx) \cdot (\nabla\varphi)(rx) r^{d-1} d\mu_1(x) \\
&= r^{d-1} \int_{\mathbb{S}^{d-1}} rx \cdot \frac{1}{r} (\nabla\varphi \circ r I_d)(x) d\mu_1(x) \\
&= r^{d-1} \int_{\mathbb{S}^{d-1}} \mathrm{Tr}(\mathrm{Hess}(\varphi \circ r I_d)(x)^t \tau_1(x)) d\mu_1(x) \\
&= r^{d-1} \int_{\mathbb{S}^{d-1}} \mathrm{Tr}(r^2 \mathrm{Hess}(\varphi)(rx)^t \tau_1(x)) d\mu_1 \\
&= \int_{\mathbb{S}_r^{d-1}} \mathrm{Tr}\left(\mathrm{Hess}(\varphi)(x)^t r^2 \tau_1\left(\frac{x}{r}\right)\right) d\mu_r(x)
\end{aligned}$$

Si bien que $\tau_r = r^2 \tau_1 \circ \frac{1}{r} I_d$ est un noyau de Stein pour μ_r, comme annoncé.

$\square$

Encore une fois, le noyau obtenu ne dépend pas de r. Dans la suite, on notera $\tau(x)$ la fonction définie sur $\mathbb{R}^d$ valant $\tau_r(x)$ sur $\mathbb{S}_r^{d-1}$.

3.3 Un noyau pour les mesures radiales

Maintenant que l'on a obtenu un noyau de Stein pour la mesure uniforme sur la sphère, on désire regarder les mesures à symétries radiales. Le fait est que les mesures μ_r partagent le même noyau vu comme une fonction sur $\mathbb{R}^d$ restreinte à $\mathbb{S}_r^{d-1}$, ce qui rend la tâche très facile.
Il s'agit en fait d'un cas particulier du problème suivant :

Etant données $(\mu_i)_i$ une famille de mesures sur $\mathbb{R}^d$, et $(a_i)_i$ une famille de réels positifs de somme 1, trouver un noyau de Stein τ de la mesure $\sum \alpha_i \mu_i$ en fonction des noyaux de Stein τ_i des μ_i. Une telle somme est appelée une mixture de mesure.
Pour le cas des mixtures de mesures uniformes, on a vu que le noyau était le même vu comme une fonction sur $\mathbb{R}^d$. Ainsi, si μ est une mixture de mesures uniformes sur le cercle et absolument continue par rapport à la mesure de Lebesgues, on peut écrire

$$d\mu = p(\|x\|)dx = p(r)S(r)dr d\mu_r$$

avec p une densité de probabilité sur $\mathbb{R}_+$ et $S(r)$ la surface de $\mathbb{S}_r^{d-1}$. Alors pour toute fonction test :

$$\begin{aligned}
\int_{\mathbb{R}^d} x \cdot \nabla\varphi(x) d\mu(x) &= \int_0^\infty \int_{\mathbb{S}_r^{d-1}} x \cdot \nabla\varphi(x) d\mu_r(x) S(r) p(r) dr \\
&= \int_0^\infty S(r)p(r) \int_{\mathbb{S}_r^{d-1}} \langle \mathrm{Hess}(\varphi)(x), \tau_r(x) \rangle d\mu_r(x) dr \\
&= \int_0^\infty \int_{\mathbb{S}_r^{d-1}} \langle \mathrm{Hess}(\varphi)(x), \tau(x) \rangle d\mu_r(x) S(r) p(r) dr \\
&= \int_{\mathbb{R}^d} \langle \mathrm{Hess}(\varphi)(x), \tau(x) \rangle d\mu(x)
\end{aligned}$$

Ainsi, τ est un noyau de Stein pour la mesure μ.

Autrement dit, on a exhibé un noyau de Stein pour toute mesure radiale absolument continue sur $\mathbb{R}^d$.
En particulier, on obtenu un noyau de Stein pour la gaussienne standard qui n'est pas l'identité (que l'on peut vérifier à la main par une intégration par parties).

3.4 Application au contrôle de la distance Wasserstein

Comme pour la section précédente, on cherche à utiliser le noyau trouvé pour obtenir une borne sur la distance de Wasserstein. On a vu dans la section précédente des exemples de mesures vérifiant $W_2(\mu_n, \gamma) \leqslant C\frac{d}{\sqrt{n}}$. L'enjeu ici n'est pas être d'améliorer le facteur $1/\sqrt{n}$ mais plutôt que le facteur d.
On commence par majorer la discrépance de Stein de μ_r.

On commence par réécrire $(d-1)\tau_r(x) = r^2 I_d - x^t x$. Cela permet notemment de remarque que τ_r est à valeurs dans l'ensemble des matrices symétriques positives puisque pour tout vecteur a et tout point x :

$$^t a \tau_r(x) a = \frac{1}{d-1}(r^{2\,t}aa - ^t a x^t x a) = \frac{1}{d-1}(||x||^2||a||^2 - \langle a, x\rangle^2) \geqslant 0$$

par Cauchy-Schwarz, avec égalité ssi a et x sont alignés avec l'origine.
On diagonalise désormais $\tau'_r := (d-1)\tau_r$.

$$\chi_{\tau'_r(x)}(X) = \det(XI_d - \tau'_r(x)) = \det\left(\left(X - r^2\right) I_d + x^t x\right) = \chi_{-x^t x}(X - r^2)$$

La matrice $-x^t x$ est de rang 1 comme produit d'une colonne et d'une ligne et est symétrique donc diagonalisable. Son polynôme caractéristique vérifie donc $\chi_{-x^t x}(Y) = Y^{d-1}(Y - \underbrace{\mathrm{Tr}(-x^t x)}_{=-r^2})$,

si bien que

$$\chi_{\tau'_r(x)}(X) = (X - r^2)^{d-1} X$$

La matrice τ'_r étant symétrique, elle est diagonalisable, on en déduit que τ_r est orthosemblable à

$$\tau_r(x) \sim \frac{1}{d-1}\begin{pmatrix} r^2 & 0 & \cdots & 0 & 0 \\ 0 & r^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & r^2 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}$$

On déduit alors que :

$$
\begin{aligned}
\|\tau(x)-I_d\|^2 &= \mathrm{Tr}((\tau(x)-I_d)(\tau(x)-I_d)) \\
&= \mathrm{Tr}(\tau(x)^2)-2\mathrm{Tr}(\tau(x))+\mathrm{Tr}(I_d) \\
&= (d-1)\frac{\|x\|^4}{(d-1)^2}-2(d-1)\frac{\|x\|^2}{d-1}+d \\
&= (d-1)\left(\frac{\|x\|^2}{d-1}-1\right)^2+1
\end{aligned}
$$

En particulier, pour la mesure μ_r, on obtient que

$$
\begin{aligned}
\int \|\tau-I_d\|^2 d\mu_r &= \int \left((d-1)\left(\frac{r^2}{d-1}-1\right)^2+1\right) d\mu_r \\
&\leqslant \left((d-1)\left(\frac{r^2}{d-1}-1\right)^2+1\right)\int d\mu_r \\
&= (d-1)\left(\frac{r^2}{d-1}-1\right)^2+1
\end{aligned}
$$

Ainsi, en prenant $r=\sqrt{d-1}$, la discrépance de μ_r ne dépend pas de la dimension. Cela correspond également au rayon de la sphère portant la plus grande masse de la gaussienne. Toutefois, on ne peut pas en tirer plus d'informations puisque la masse se répartissant sur les autres sphères n'est pas négligeable pour la gaussienne.
En revanche, pour les mesures admettant un moment d'ordre 4, on a le contrôle suivant :

$$
\begin{aligned}
S(\mu)^2 &\leqslant \int (d-1)\frac{r^4}{(d-1)^2}p(r)\mu_r dr + d\int p(r)\mu_r dr - 2\int r^2 p(r)\mu_r dr \\
&= d-2\int \|x\|^2 d\mu + \frac{1}{d-1}\int \|x\|^4 d\mu
\end{aligned}
$$

On récupère donc un facteur d pour $S(\mu)^2$, et donc un facteur $\sqrt{d}$ pour la distance de Wasserstein, ce qui est connu comme la meilleure approximation possible.

3.5 Projection de la mesure uniforme sur la première coordonnée

Le noyau de Stein établi précédemment permet de retrouver un résultat établi dans [6] par Meckes et Meckes. Même si cela n'est pas formulé explicitement en terme de noyaux de Stein dans l'article, il s'agit de déterminer un noyau de Stein pour la mesure donnant la loi de la première coordonnée d'une variable aléatoire à symétrie sphérique.
Le résultat annoncé est le suivant : si $X=(X_1,\dots,X_n)$ est une variable aléatoire à symétrie sphérique, alors $E[X_2^2|X_1]$ est un noyau de Stein pour X_1.
Notre expression de τ permet de retrouver presque directement ce résultat, par l'intermédiaire du lemme suivant :

Lemme 3.1. *Soit $X=(X_1,\dots,X_2)$ une variable aléatoire de loi μ. On suppose que μ admet un noyau de Stein τ. Alors $\tau^i(x_i) := E[\tau_{i,i}(X)|X_i=x_i]$ est un noyau de Stein pour X_i.*

Démonstration. Soit f une fonction de la variable x_i. On pose ensuite $g = (0, \dots, 0, f, 0 \dots 0)$ (avec f en i-ème position).

$$\begin{aligned} E[\tau_{i,i}(X) f''(X_1)] &= E[\langle \tau(X), \mathrm{Hess} g(X) \rangle] \\ &= E[X \cdot \nabla g(X)] \\ &= E[X_i f'(X_i)] \end{aligned}$$

ce qui fournit le résultat voulu. □

Pour voir ensuite que $E[X_2^2|X_1]$ est bien un noyau de Stein pour X_1, il suffit de remarquer que

$$E[X_2^2|X_1] = \frac{1}{d-1} E\left[\sum_{i=1}^{d} X_i^2 - X_1^2 | X_1\right] = \frac{1}{d-1} E\left[||X||^2 - X_1^2 | X_1\right]$$

et on retrouve bien en $\dfrac{||x||^2 - x_1^2}{d-1}$ la coordonnée $\tau_{1,1}$ du noyau de Stein de la mesure uniforme.

4 Noyau de Stein d'ordre supérieur pour les fonctions de Gaussiennes

Dans cette section, on cherche un noyau de Stein pour les variables aléatoires de la forme $g(X)$, avec X une gaussienne standard, avec g vérifiant de bonnes hypothèses.
Grâce au noyau trouvé pour la mesure radiale, nous sommes en mesure d'assurer l'existence d'un tel noyau. Toutefois, cela implique de déterminer la densité de la variable $g(X)$, ce qui n'est pas très pratique. On cherche donc un noyau dont l'expression dépend de la fonction g uniquement. C'est ce qui est fait par Chatterjee dans [7] et dont nous présentons la preuve dans la première sous-section.
L'enjeu est ensuite de chercher des noyaux de Stein d'ordre supérieur pour $g(X)$. L'intérêt est qu'en rajoutant des conditions pour g, notamment en rajoutant de la régularité et en supposant que les premiers moments de $g(X)$ coïncident avec ceux de X, on peut trouver des fonctions permettant de réaliser une intégration par partie supplémentaire. Si les premiers moments coïncident, on peut alors s'attendre à ce que la mesure de $g(X)$ soit proche de celle de X. On espère donc ensuite, via la méthode de Stein, obtenir une vitesse de convergence plus rapide pour $W_2(\mu_n, \gamma)$.

4.1 Cas du noyau d'ordre 1

Le noyau de Stein obtenu par Chatterjee est le suivant :

Théorème 4.1. *Soit X une gaussienne standard, g une fonction de classe C^∞ de $\mathbb{R}^n$ dans $\mathbb{R}$ et $W = g(X)$. On suppose que $E(W) = 0$ et $E(W^2) = 1$. Alors la fonction $\bar{\tau}_1$ définie par $\bar{\tau}_1(w) = E[T_1(X)|W = w]$, où*

$$T_1(x) = \int_0^{\pi/2} \cos(t) E\left[\sum_{i=1}^{n} \partial_i g(x) \partial_i g(\sin(t)x + \cos(t)X)\right] dt$$

vérifie $\mathsf{E}[f(W)W] = \mathsf{E}[f'(W)\bar{\tau}_1(W)]$ *pour toute fonction lipschitzienne* f. *Il s'agit donc d'un noyau de Stein pour* W.

Remarque : La classe de fonctions considérées n'est pas la même que celles choisies auparavant. Notons que le théorème de Rademacher assure que les fonctions f choisies sont dérivables presque partout.

Démonstration. Il suffit de démontrer que pour toute fonction lipschitizienne f, $\mathsf{E}[f(W)W] = \mathsf{E}[f'(W)T_1(X)]$.
La clé est d'introduire une copie X' de X indépendante de X et d'utiliser l'interpolation

$$W_t = g\left(\sin(t)X + \cos(t)X'\right)$$

. On peut alors écrire

$$\begin{aligned}
\mathsf{E}[f(W)W] &= \mathsf{E}[f(W)W] - \mathsf{E}[f(W)]\mathsf{E}[g(X')] \\
&= \mathsf{E}[f(W)(W_{\pi/2} - W_0)] \\
&= \mathsf{E}\left[f(W)\int_0^{\pi/2} \frac{\partial W_t}{\partial t} dt\right] \\
&= \mathsf{E}\left[f(W)\int_0^{\pi/2} \sum_{i=1}^{n} \left(\cos(t)X - \sin(t)X'\right)\partial_i g\left(\sin(t)X + \cos(t)X'\right) dt\right]
\end{aligned}$$

On pose alors $V_t = \cos(t)X - \sin(t)X'$ et $U_t = \sin(t)X + \cos(t)X'$. On a alors $X = \cos(t)V_t + \sin(t)U_t$. Les variables U_t et V_t sont deux gaussiennes standards indépendantes. On a alors :

$$\begin{aligned}
&\mathsf{E}[f(W)W] \\
= &\int_0^{\pi/2} \sum_{i=1}^{n} \mathsf{E}\left[V_t \partial_i g(U_t) f(g(\cos(t)V_t + \sin(t)U_t))\right] dt \\
= &\int_0^{\pi/2} \sum_{i=1}^{n} \int \partial_i g(u_t) e^{-u_t^2/2} \int v_t^i f(g(\cos(t)v_t + \sin(t)u_t)) e^{-v_t^{i2}/2} dv_t du_t dt \\
= &\int_0^{\pi/2} \sum_{i=1}^{n} \int \partial_i g(u_t) e^{-u_t^2/2} \int \cos(t)\partial_i g(\cos(t)v_t + \sin(t)u_t) f'(g(\cos(t)v_t + \sin(t)u_t)) e^{-v_t^{i2}/2} dv_t du_t dt \\
= &\int_0^{\pi/2} \cos(t) \sum_{i=1}^{n} \mathsf{E}[\partial_i g(X)\partial_i g(\sin(t)X + \cos(t)X') f'(W)] dt \\
= &\ \mathsf{E}[f'(W)T_1(X)]
\end{aligned}$$

sous réserve de justification des différentes intégrations, qui se font en utilisant que f est Lipschitizienne. $\square$

On se permet ici une petite digression autour de la preuve présentée. L'article [7] de Chatterjee n'utilisait pas la même interpolation pour W_t, puisqu'il utilisait plutôt l'interpolation suivante entre 0 et 1 :

$$W_t = g\left(\sqrt{t}X + \sqrt{1-t}X'\right)$$

Un changement de variable $e^{-2u} = t$ permet d'obtenir l'interpolation entre 0 et $+\infty$:

$$W_u = g\left(e^{-u}X + \sqrt{1-e^{-2u}}X'\right)$$

C'est alors qu'on retrouve ici la famille de variables aléatoires du sous-groupe d'Ornstein-Uhlenbeck, qui est le sous-groupe $(P_t g)$ engendré par le générateur infinitésimal Hess $g - x \cdot \nabla g$. Cette version permet donc de mieux motiver l'interpolation de la preuve, au delà du besoin de générer des Gaussiennes indépendantes, puisqu'on retrouve une famille dont les propriétés sont bien connues dans le domaine.

Toutefois, l'inconvénient de l'usage de $\sqrt{t}$ est qu'il ne passe pas pour une intégration par parties supplémentaire (qui est ce que nous souhaitons faire dans la suite) puisque la racine n'est pas dérivable en 0. Le choix de $\sin(t)$ et $\cos(t)$ est donc un choix personnel de confort, qui permet de toujours garder des gaussiennes orthogonales à tout t et sans se soucier des termes de bords des intégrales, même s'il occulte la motivation derrière l'opération.

4.2 Cas du noyau d'ordre 2

Le but de cette sous-section est d'adapter la preuve précédente pour obtenir des noyaux de Stein d'ordre supérieur.

Définition 4.1. *Soit μ une mesure. On dit que $\bar{\tau}_n$ est un noyau de Stein de dimension n pour μ si $\bar{\tau}_n$ vérifie la formule d'intégration par partie suivante*

$$E(f^{(n)}(W)\bar{\tau}_n(W)) = E(Wf(W) - f'(W))$$

pour toute fonction test f.

Dans la suite, on notera également τ_n un noyau vérifiant la formule d'intégration par partie suivante :

$$E[f(W)H_n(W)] = E[f'(W)\tau_n(W)]$$

pour toute fonction test f , avec H_n le $n-$ème polynôme de Hermitte et un tel noyau sera dit d'ordre n.

La motivation derrière l'introduction de ces objets est toujours la même : regarder quelles formules d'intégration par parties vérifie la gaussienne puis comparer avec la formule vérifiée par la mesure étudiée. On note que le noyau de dimension 1, s'il existe, correspond à $\tau - I_d$, où τ est le noyau avec lequel on travaille usuellement.

La comparaison change un peu, puisqu'au lieu de comparer avec la fonction identité, on compare les noyaux de dimension supérieure avec 0 et on compare les noyaux d'ordre n avec le $n-1$-ème polynôme de Hermitte.

Dans la suite, sauf mention du contraire, on ne s'intéressera qu'aux noyaux de dimension 2 ou aux noyaux d'ordre 2. On cherche ici un noyau de dimension 2 pour $g(X)$. La démonstration ci-dessus ne s'adaptant pas directement pour obtenir un tel noyau, on va d'abord déterminer un noyau d'ordre 2, et utiliser une relation entre ordre et dimension. Pour le cas $n = 2$, on a le résultat suivant :

Lemme 4.1. *Si τ_1 et τ_2 sont des noyaux respectivement d'ordre 1 et 2 pour μ pour une certaine classe de fonctions stable par dérivation et multiplication par un polynôme, alors $\bar{\tau}_2 := \tau_2 - W\tau_1$ est un noyau de dimension 2 pour μ,*

Démonstration. On calcule $E[f''(W)\bar{\tau}_2(W)]$ pour une certaine classe de fonctions.

$$\begin{aligned}
E[f''(W)\bar{\tau}_2(W)] &= E[f''(W)\tau_2(W)] - E[f''(W)W\tau_1(W)] \\
&= E[f'(W)(W^2-1)] - E[f''(W)W\tau_1(W)] \\
&= E[f'(W)(W^2-1)] + E[f'(W)\tau_1] - E[(f''(W)W + f'(W))\tau_1(W)] \\
&= E[f'(W)(W^2-1)] + E[Wf(W)] - E[W^2 f'(W)] \\
&= E[Wf(W)] - E[f'(W)]
\end{aligned}$$

ce qui donne le résultat. □

On cherche donc ici à calculer τ_2. Cette fois-ci, la démonstration par interpolation s'adapte bien, et on a presque instantanément le résultat suivant pour calculer τ_n pour tout n :

Théorème 4.2. *Soit X une gaussienne standard, g une fonction de classe C^∞ et $W = g(X)$. On suppose que $E(W^k) = E(X^k)$ pour tout $k \leqslant n+1$. On note H_n le $n-$ème polynôme de Hermite. Alors la fonction τ_n définie par $\tau_n(w) = E[T_n(X)|W = w]$, où*

$$T_n(x) = n\int_0^{\pi/2} \cos(t) \sum_{i=1}^{n} E[\partial_i g(x)\partial_i g(\sin(t)x + \cos(t)X)H_{n-1}(g(\sin(t)x + \cos(t)X))]dt$$

vérifie $E[f(W)H_n(W)] = E[f'(W)\tau_n(W)]$ pour toute fonction lipschitzienne.

Démonstration. Commençons par rappeler que $H_n' = nH_{n-1}$. De plus, la famille (H_n) étant orthogonale pour $L^2(\mathbb{R}, e^{-x^2/2}dx)$, on a pour tout $n \geqslant 1$, $E[H_n(W)] = E[H_n(X)] = \langle H_n, H_0\rangle = 0$. Ainsi

$$\begin{aligned}
E[f(W)H_n(W)] &= E[f(W)(H_n(W_{\pi/2}) - H_n(W_0))] \\
&= E\left[f(W)\int_0^{\pi/2} \partial_t H_n(W_t)dt\right] \\
&= E\left[f(W)n\int_0^{\pi/2} \sum_{i=0}^{n} V_t^i \partial_i g(U_t)H_{n-1}(g(U_t))dt\right] \\
&= n\int_0^{\pi/2} \sum_{i=0}^{n} E[V_t^i \partial_i g(U_t)H_{n-1}(g(U_t))f(g(\sin(t)U_t + \cos(t)V_t))]dt \\
&= n\int_0^{\pi/2} \sum_{i=0}^{n} E[\partial_i g(U_t)H_{n-1}(g(U_t))\cos(t)\partial_i g(X)f'(g(X))]dt \\
&= E[f'(W)T_n(X)]
\end{aligned}$$

ce qui nous permet de conclure. □

On peut alors déterminer $\bar{\tau}_2$. En particulier, en posant $T := T_2 - WT_1$, $\bar{\tau}_2 := E[T(X)|W]$ est un noyau de Stein de dimension 2. On a alors

$$T(x) = \int_0^{\pi/2} \cos(t) \sum_{i=1}^{n} E[\partial_i g(X) \partial_i g(U_t)(2g(U_t) - g(X))] dt$$

4.3 Comparaison avec la convergence obtenue pour le noyau d'ordre 1

Dans [7], Chatterjee étudie les matrices de Wigner :

Définition 4.2. *On appelle matrice de Wigner une matrice symétrique aléatoire dont les coefficients sont des variables aléatoires indépendantes* $(X_{i,j})_{1 \leqslant i,j \leqslant n}$ *(avec* $X_{i,j} = X_{j,i}$*).*

Plus précisément, Chatterjee s'intéresse aux matrices A_n dont les coefficients sont de la forme $\frac{1}{\sqrt{n}} X_{i,j}$ avec $X_{i,j}$ une fonction de gaussienne ayant les mêmes premiers moments. En posant $W_n = \mathrm{Tr}(f(A_n))$ pour f une fonction entière avec de bonnes propriétés, il s'agit d'estimer la vitesse de convergence de $\frac{W_n - E[W_n]}{\sqrt{\mathrm{Var}(W_n)}}$ vers une gaussienne standard.
L'étude nous ramène à estimer la discrépance de Stein d'une variable de la forme $g(X)$ avec X une gaussienne. La suite de notre travail a consisté à reprendre la méthode employée par Chatterjee et l'adapter au noyau de Stein de dimension supérieure, et c'est ce que nous présentons ici.

Dans la suite, on considère n fonctions $\varphi_1, \ldots, \varphi_n$ de $\mathbb{R}$ dans $\mathbb{R}$, n gaussiennes standard indépendantes $X_1, \ldots, X_n$ ainsi que les variables $Y_i = \varphi_i(X_i)$ et une fonction $f : \mathbb{R}^d \to \mathbb{R}$. La fonction $g := f \circ (\varphi_1, \ldots, \varphi_n)$, que l'on pourra aussi noter $f \circ \varphi$ permet de définir une variable aléatoire $g(X)$ qui est une fonction d'une gaussienne. On pose alors

$$\begin{aligned}
\kappa_0 &= E\left[\left(\sum_{i=1}^{n} |\partial_i f(X)|^4\right)^2\right]^{1/4} \\
\kappa_1 &= E\left[|\nabla f|^8\right]^{1/8} \\
\kappa_2 &= E\left[||\mathrm{Hess}\ f||^8\right]^{1/8}
\end{aligned}$$

On suppose que κ_0, κ_1 et κ_2 sont finis. On suppose également que les φ_i vérifient $||\varphi_i'||_\infty \leqslant c_1$ et $||\varphi_i''||_\infty \leqslant c_2$ pour c_1 et c_2 des constantes.
A titre d'exemple, on peut considérer le cas où les fonctions φ_i sont l'identité et $f(x) = \sum x_i x_{i+1}/\sqrt{n}$. On a alors

$$\partial_i g(x) = \frac{x_{i-1} + x_{i+1}}{\sqrt{n}}$$

et

$$\partial^2_{i,j} = \frac{1}{\sqrt{n}} 1_{|i-1|=1}$$

Dans ce cas, on a $\kappa_0 = O(1/\sqrt{n})$, $\kappa_1 = O(1)$ et $\kappa_2 = O(1/\sqrt{n})$. Pour cette dernière estimation, on utilise que le rayon spectrale d'une matrice tridiagonale ne dépend pas de la taille de la matrice.

On peut désormais attaquer l'étude.

Comme annoncé, on ne doit pas comparer $\bar{\tau}_2$ avec 1, mais bien avec 0, car c'est la fonction nulle qui satisfait la formule d'intégration par parties étudiée. On veut donc estimer $\int |\bar{\tau}_2|^2 d\mu = \mathsf{E}[|\mathsf{E}[\mathsf{T}(X)|W]|^2]$. Puisque l'espérance conditionnelle est une projection orthogonale, on a par Pythagore :

$$\mathsf{E}[|\mathsf{E}[\mathsf{T}(X)|W]|^2] \leqslant \mathsf{E}[\mathsf{T}(X)^2] = Var[\mathsf{T}(X)]$$

On cherche dans la suite à majorer $Var[\mathsf{T}(X)]$.
D'après l'inégalité de Poincaré pour la Gaussienne :

$$Var[\mathsf{T}(X)] \leqslant \mathsf{E}[|\nabla \mathsf{T}|^2]$$

Or on a

$$\begin{aligned}
\nabla \left(\textstyle\sum_i \partial_i g(x) \partial_i g(u_t)(2g(u_t) - g(x)) \right) &= \nabla \left({}^t\nabla g(x) \nabla g(u_t)(2g(u_t) - g(x)) \right) \\
&= \nabla g(x) \cdot \nabla g(u_t)(2\sin(t)\nabla g(u_t) - \nabla g(x)) \\
&+ \sin(t)(2g(u_t) - g(x)) \text{Hess } g(u_t) \nabla g(x) \\
&+ (2g(u_t) - g(x)) \text{Hess } g(x) \nabla g(u_t)
\end{aligned}$$

On utilise ensuite l'inégalité triangulaire, l'inégalité de Cauchy-Schwarz pour les intégrales ainsi que l'inégalité $(a+b+c)^2 \leqslant 3(a^2+b^2+c^2)$. On a alors

$$\begin{aligned}
\mathsf{E}[|\nabla \mathsf{T}|^2] &\leqslant 3\frac{\pi}{2} \int_0^{\pi/2} \cos^2(t) \mathsf{E}\left[|\nabla g(X) \cdot \nabla g(U_t) \cdot (2\sin(t)\nabla g(U_t) - \nabla g(X))|^2 \right] dt \quad (1) \\
&+ 3\frac{\pi}{2} \int_0^{\pi/2} \cos^2(t) \mathsf{E}\left[(2g(U_t) - g(X))^2 |\text{Hess } g(U_t) \nabla g(X)|^2 \right] dt \quad (2) \\
&+ 3\frac{\pi}{2} \int_0^{\pi/2} \cos^2(t) \sin^2(t) \mathsf{E}\left[(2g(U_t) - g(X))^2 \|\text{Hess } g(X) \nabla g(U_t)\|^2 \right] dt \quad (3)
\end{aligned}$$

On estime ensuite séparément (1), (2) et (3).

On commence par l'étude de (2).
Tout d'abord, en notant $\varphi(y) = (\varphi_1(y_1), \ldots, \varphi_n(y_n))$

$$\partial_i g(y) = \varphi_i'(y_i) \partial_i f(\varphi(y))$$

Puis, si $i \neq j$,

$$\partial_{i,j} g = \varphi_i'(y_i) \varphi_j'(y_j) \partial^2_{i,j} f(\varphi(y))$$

et

$$\partial_{i,i} g = (\varphi_i'(y_i))^2 \partial^2_{i,i} f(\varphi(y)) + \varphi_i''(y_i) \partial_i f(\varphi(y))$$

On a ensuite, toujours par l'inégalité $(a+b)^2 \leqslant 2(a^2+b^2)$:

$$
\begin{aligned}
|\text{Hess } g(u_t)\nabla g(x)|^2 &= \sum_{i=1}^{n}\left(\sum_{j=1}^{n}\partial^2_{i,j} g(u_t)\partial_j g(x)\right)^2 \\
&\leqslant 2\sum_{i=1}^{n}\left(\sum_{j=1}^{n}\varphi_i'(u_t^i)\varphi_j'(u_t^j)\partial^2_{i,j} f(\varphi(u_t))\varphi_j'(x_j)\partial_j f(\varphi(x))\right)^2 \\
&+ 2\sum_{i=1}^{n}\left(\varphi_i''(u_t^i)\partial_i f(\varphi(u_t))\varphi_i'(x_i)\partial_i f(\varphi(x))\right)^2 \\
&\leqslant 2c_1^6|\text{Hess} f(\varphi(u_t))\nabla f(\varphi(x))|^2 + 2c_2^2c_1^2\sum_{i=1}^{n}(\partial_i f(\varphi(u_t))\partial_i f(\varphi(x)))^2
\end{aligned}
$$

Notons ensuite que $E[(2g(U_t) - g(X))^4]$ est majorée par une constante C qui ne dépend pas de n puisque $E[g(X)^4] = E[X^4]$ et que X et U_t sont des gaussiennes. Ensuite, l'inégalité de Cauchy-Schwarz et les inégalités $(a+b)^2 \leqslant 2(a^2+b^2)$ et $\sqrt{a+b} \leqslant \sqrt{a} + \sqrt{b}$ donnent :

$$
\begin{aligned}
& E[\left[(2g(U_t) - g(X))^2\,|\text{Hess } g(U_t)\nabla g(X)|^2\right] \\
\leqslant\ & E[(2g(U_t) - g(X))^4]^{1/2} E\left[|\text{Hess } g(U_t)\nabla g(X)|^4\right]^{1/2} \\
\leqslant\ & C^{1/2} E\left[\left(2c_1^6|\text{Hess } f(\varphi(U_t))\nabla f(\varphi(X))|^2 + 2c_2^2c_1^2\sum_{i=1}^{n}(\partial_i f(\varphi(U_t))\partial_i f(\varphi(X)))^2\right)^2\right]^{1/2} \\
\leqslant\ & 2\sqrt{2}C^{1/2}\left(E\left[c_1^{12}\,|\text{Hess } f(\varphi(U_t))\nabla f(\varphi(X))|^4\right]^{1/2} + E\left[c_1^4c_2^4\left(\sum_{i=1}^{n}(\partial_i f(\varphi(U_t))\partial_i f(\varphi(X)))^2\right)^2\right]^{1/2}\right)
\end{aligned}
$$

On traite alors séparément les deux termes de la parenthèse.
D'une part, d'après l'inégalité de Cauchy-Schwarz :

$$
\begin{aligned}
& E\left[|\text{Hess } f(\varphi(U_t))\nabla f(\varphi(X))|^4\right] \\
\leqslant\ & E[\|\text{Hess } f(\varphi(U_t))\|^8]^{1/2} E[|\nabla f(\varphi(X))|^8]^{1/2} \\
=\ & E[\|\text{Hess } f(\varphi(X))\|^8]^{1/2} E[|\nabla f(\varphi(X))|^8]^{1/2} \\
=\ & \kappa_2^4\kappa_1^4
\end{aligned}
$$

D'autre part, toujours d'après Cauchy-Schwarz :

$$
\left(\sum_{i=1}^{n}(\partial_i f(\varphi(u_t))\partial_i f(\varphi(x)))^2\right)^2 \leqslant \left(\sum_{i=1}^{n}\partial_i f(\varphi(u_t))^4\right)\left(\sum_{i=1}^{n}\partial_i f(\varphi(x))^4\right)
$$

si bien que

$$\begin{aligned}
& \mathsf{E}\left[\left(\sum_{i=1}^{n}(\partial_i f(\varphi(U_t))\partial_i f(\varphi(X)))^2\right)^2\right] \\
\leqslant\ & \mathsf{E}\left[\left(\sum_{i=1}^{n}\partial_i f(\varphi(U_t))^4\right)\left(\sum_{i=1}^{n}\partial_i f(\varphi(X))^4\right)\right] \\
\leqslant\ & \mathsf{E}\left[\left(\sum_{i=1}^{n}\partial_i f(\varphi(U_t))^4\right)^2\right]^{1/2}\mathsf{E}\left[\left(\sum_{i=1}^{n}\partial_i f(\varphi(X))^4\right)^2\right]^{1/2} \\
=\ & \kappa_0^4
\end{aligned}$$

Au total, on obtient que

$$\mathsf{E}\left[(2g(U_t)-g(X))^2|\text{Hess } g(U_t)\nabla g(X)|^2\right] \leqslant 2\sqrt{2}C^{1/2}(c_1^6\kappa_1^2\kappa_2^2 + c_1^2c_2^2\kappa_0^2)$$

On obtient de la même façon que

$$\mathsf{E}\left[(2g(U_t)-g(X))^2|\text{Hess } g(X)\nabla g(U_t)|^2\right] \leqslant 2\sqrt{2}C^{1/2}(c_1^6\kappa_1^2\kappa_2^2 + c_1^2c_2^2\kappa_0^2)$$

Il reste alors à contrôler

$$\mathsf{E}\left[|\nabla g(X)\cdot\nabla g(U_t)(2\sin(t)\nabla g(U_t)-\nabla g(X))|^2\right]$$

Avec plusieurs itérations de Cauchy-Schwarz, on trouve

$$\begin{aligned}
& \mathsf{E}\left[|\nabla g(X)\cdot\nabla g(U_t)\cdot(2\sin(t)\nabla g(U_t)-\nabla g(X))|^2\right] \\
\leqslant\ & \mathsf{E}[|\nabla g(X)|^8]^{1/4}\mathsf{E}[|\nabla g(U_t)|^8]^{1/4}\mathsf{E}[|2\sin(t)\nabla g(U_t)-\nabla g(X)|^4]^{1/2} \\
\leqslant\ & c_1^4\kappa_1^4\mathsf{E}[|2\sin(t)\nabla g(U_t)-\nabla g(X)|^4]^{1/2}
\end{aligned}$$

On utilise ensuite que $|a-b|^4 \leqslant (|a|+|b|)^4 \leqslant 8(|a|^4+|b|^4)$ pour avoir que :

$$\mathsf{E}[|2\sin(t)\nabla g(U_t)-\nabla g(X)|^4] \leqslant 8\mathsf{E}[16\sin(t)^4|\nabla g(U_t)|^4] + 8\mathsf{E}[|\nabla g(X)|^4] \leqslant 256c_1^4\kappa_1^4$$

Finalement, $\mathsf{E}\left[|\nabla g(x)\cdot\nabla g(u_t)\cdot(2\sin(t)\nabla g(u_t)-\nabla g(x))|^2\right] \leqslant 16c_1^6\kappa_1^6$.
Au total, on obtient une estimation de

$$Var[T(X)] \leqslant C(c_1^6\kappa_1^2\kappa_2^2 + c_1^2c_2^2\kappa_0^2 + c_1^6\kappa_1^6)$$

Malheureusement, cette estimation est en réalité moins efficace que celle obtenue par Chatterjee ! En effet, celui-ci, avec le noyau d'ordre 1, obtient une estimation de.

$$Var[T_1(X)] \leqslant C(c_1^6\kappa_1^2\kappa_2^2 + c_1^2c_2^2\kappa_0^2)$$

On voit avec la fonction $x \mapsto \sum x_i x_{i+1}/\sqrt{n}$ que notre estimation n'est plus qu'un $O(1)$ tandis que celle de [7] est un $O(1/\sqrt{n})$ si on applique à la distance de Wasserstein.

Par manque de temps, nous n'avons pas pu examiner de plus près ce qui fait échouer la méthode, pourtant très prometteuse.

Références

[1] C. STEIN, A Bound for the Error in the Normal Approximation to the Distribution of a Sum of Dependant Random Variable, 1972

[2] T. COURTADE, M.FATHI, A.PANAJADI, Existence of Stein Kernel under a Spectral Gap, and Discrepancy Bounds, 2017

[3] A.SAUMARD, Weighted Poincaré inequalities, concentration inequalities and tail bounds related to Stein kernels in dimension one, 2018

[4] C.ANE, S.BLACHERE, D.CHAFAI, P.FOUGERES, I.GENTIL, F.MALRIEU, C.ROBERTO, G.SCHEFFER, Sur les inégalités de Sobolev logarithmiques, 2001

[5] M.FATHI, Stein kernels and moment maps, 2018

[6] E.MECKES, M.MECKES, The Central Limit Problem for Random Vectors with Symmetries, 2007

[7] S.CHATTERJEE, Fluctuations of eigenvaluesand second order Poincaré inequalities, 2007

[8] D.CORDERO-ERAUSQUIN, B.KLARTAG, Moment measures, 2015

[9] R.BERMAN, B.BERNDTSSON, Real Monge-Amper equations and Kähler-Ricci solitons on toric log Fano varieties, 2013

[10] B.KLARTAG, Logarithmically-concave moment measures I. Geometric Aspects of Functional Analysis, 2014