Operations with Concentration Inequalities

Cosme Louart

Assistant Professor at CUHK Shenzhen

SCHOOL OF DATA SCIENCE

Content

I - Motivation in machine learning

Given a data set $\mathcal{D} = ((x_1, y_1), \dots, (x_n, y_n))$ n independent drawings of $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}$

Look for a mapping $\Phi_{\mathcal{D}} : \mathbb{R}^p \to \mathbb{R}$ that "minimizes":

$$L(\Phi_{\mathcal{D}}(X), Y) \qquad \text{For a given a loss } L : \mathbb{R}^2 \to \mathbb{R}_+$$

Behavior of the loss $L(\Phi_{\mathcal{D}}(X), Y)$?

Ex: $\mathcal{D}, X \to \Phi_{\mathcal{D}}(X)$ $\lambda_{n,p}$-Lipschitz:

$$\mathbb{P}\left(|\Phi_{\mathcal{D}}(X) - \Phi_{\mathcal{D}'}(X')| \geq t\right) \leq \alpha\left(\frac{t}{\lambda_{np}}\right)$$

Idea: Consider $\alpha : t \mapsto \sup\left\{\mathbb{P}\left(|f(X) - f(X')| \geq t\right),\ f : \mathbb{R}^p \to \mathbb{R}, \text{1-Lipschitz}\right\}$.

Question: $\lim_{t\to\infty} \alpha(t) = 0$? Dependence on p, n? When Φ non Lipschitz ?

I - Motivation in machine learning

Theorem: Given $Z \sim \mathcal{N}(\mu, I_n)$, $\forall f : \mathbb{R}^n \to \mathbb{R}$, 1-Lipschitz:

$$\mathbb{P}\left(|f(Z) - f(Z')| \geq t\right) \leq 2e^{-\frac{t^2}{2}} \quad Z, Z' \; i.i.d.$$

Given $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ λ-Lipschitz and $f : \mathbb{R}^n \to \mathbb{R}$ 1-Lipschitz:

$$\mathbb{P}\left(|f(\Phi(Z)) - f(\Phi(Z'))| \geq t\right)$$

$$= \mathbb{P}\left(\left|\frac{1}{\lambda} f(\Phi(Z)) - \frac{1}{\lambda} f(\Phi(Z'))\right| \geq \frac{t}{\lambda}\right) \leq 2e^{-\frac{t^2}{2\lambda^2}}.$$

Insight in section 2:

$$\|\Phi(Z) - \Phi(Z')\| \leq \Lambda \|Z - Z'\| \quad a.s.$$

Random

Theorem: (Talagrand)
Given $Z = (Z_1, \ldots Z_n) \in [0,1]^n$ s.t. $Z_1, \ldots, Z_n$ independent
$\forall f : \mathbb{R}^p \to \mathbb{R}$, 1-Lipschitz and convex:

$$\mathbb{P}\left(|f(Z) - \mathbb{E}[f(Z)]| \geq t\right) \leq 2e^{-\frac{t^2}{4}}.$$

Michel Talagrand (1995) *Concentration of measure and isoperimetric inequalities in product spaces.* Publications mathématiques de l'IHÉS, 104:905–909.

I - Motivation in machine learning

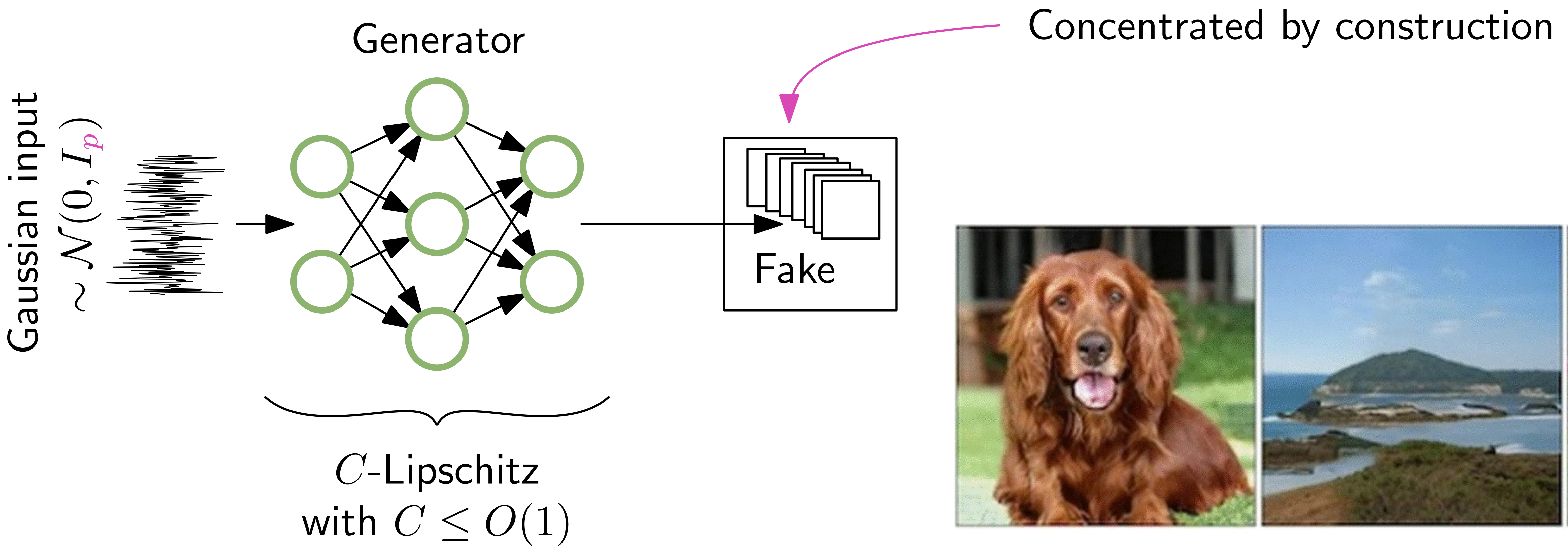
Theorem: Given $Z \sim \mathcal{N}(\mu, I_n)$, $\forall f : \mathbb{R}^n \to \mathbb{R}$, 1-Lipschitz:

$$\mathbb{P}\left(|f(Z) - \mathbb{E}[f(Z)]| \geq t\right) \leq 2e^{-\frac{1}{2}t^2}$$

Recall: $\forall \Phi : \mathbb{R}^n \to \mathbb{R}^q$ λ-Lipschitz, $\forall f : \mathbb{R}^n \to \mathbb{R}$, 1-Lipschitz:

$$\mathbb{P}\left(|f(\Phi(Z)) - \mathbb{E}[f(\Phi(Z))]| \geq t\right) \leq 2e^{-\frac{1}{2}(t/\lambda)^2}.$$

GAN generated images are concentrated vectors



I - Motivation in machine learning

Outside from Gaussian contration:
Possible to set heavy tailed concentration depending on the dimension

Proposition:

Consider $X = (X_1, \dots, X_n) \in \mathbb{R}^n$ and $Z = (Z_1, \dots, Z_n) \sim \mathcal{N}(0, I_n)$ such that:

- $\forall i \in [n] : \ X_i = \phi_i(Z_i)$
- $\exists C, q > 0, \ \forall t \in \mathbb{R}, \forall i \in [n]$: $|\phi_i'(t)| \leq \frac{C}{|t|} \exp(\frac{t^2}{2q})$

$$\Longrightarrow \quad \mathbb{E}[|X_i|^r] = \mathbb{E}[|\phi_i(Z_i)|^r] \leq \mathbb{E}[|Z_i \phi_i'(Z_i)|^r] \leq C' \int z e^{-\frac{z^2}{2}(1-\frac{r}{q})} dz$$

Then For all $f : \mathbb{R}^n \mapsto \mathbb{R}$ 1-Lipschitz:

$$\mathbb{P}\left(|f(X) - f(X')| \geq t\right) \leq \frac{Cn}{t^q}$$

Dependence on the dim:
Stand. dev. $\sim n^{\frac{1}{q}}$

NB: $\forall r < q : \mathbb{E}[|X_i|^r] < \infty$ and $\mathbb{E}[|f(X)|^r] < \infty$

II - Operation with concentration inequalities.

Definition: $\alpha \boxplus \beta = (\alpha^{-1} + \beta^{-1})^{-1}$

Proposition: Given $\alpha, \beta : \mathbb{R}_+ \to \mathbb{R}_+$, two random variables $X, Y \in \mathbb{R}$ such that $\forall t \in \mathbb{R}$:

$$\mathbb{P}\left(X \geq t\right) \leq \alpha(t) \quad \text{and} \quad \mathbb{P}\left(Y \geq t\right) \leq \beta(t)$$

Then $\mathbb{P}\left(X + Y \geq t\right) \leq 2\alpha \boxplus \beta(t)$

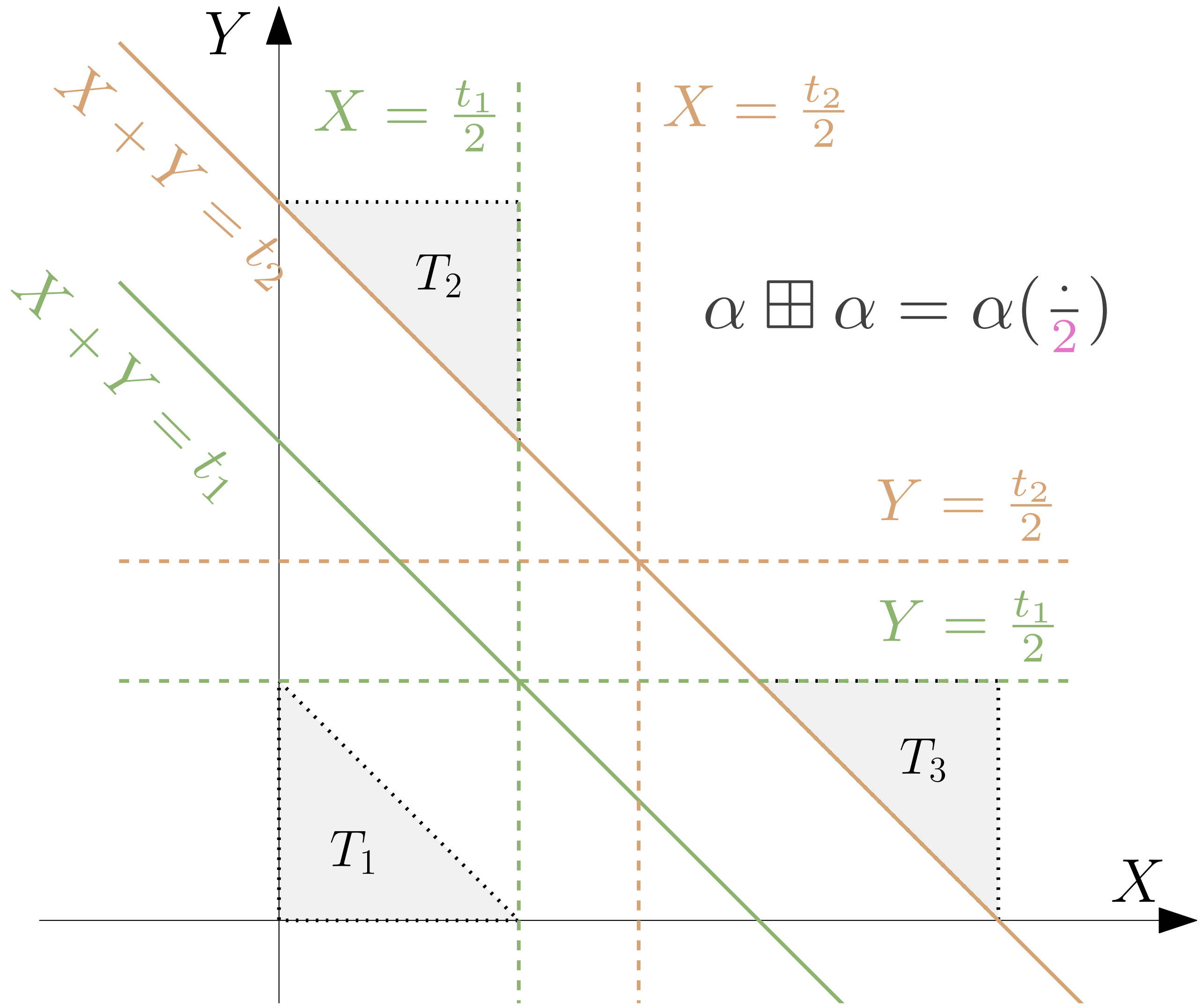
Proof: Denoting $\gamma \equiv \alpha \boxplus \beta$, for any $t \in \mathbb{R}$:

In particular: $\alpha^{-1}(\gamma(t)) + \beta^{-1}(\gamma(t)) = t$

$$\begin{aligned}\mathbb{P}\left(X + Y \geq t\right) &\leq \mathbb{P}\left(X + Y \geq \alpha^{-1}(\gamma(t)) + \beta^{-1}(\gamma(t))\right)\\ &\leq \mathbb{P}\left(X \geq \alpha^{-1}(\gamma(t))\right) + \mathbb{P}\left(Y \geq \beta^{-1}(\gamma(t))\right)\\ &\leq 2\gamma(t)\end{aligned}$$

$\forall t \in [t_1, t_2]$:

$$\mathbb{P}(X + Y \geq t) = \tfrac{2}{3} = \mathbb{P}(X \geq \tfrac{t}{2}) + \mathbb{P}(Y \geq \tfrac{t}{2})$$



Uniform distribution of (X, Y) on T_1, T_2, T_3

II - Operation with concentration inequalities.

Definition: $\alpha \boxtimes \beta \equiv (\alpha^{-1} \cdot \beta^{-1})^{-1}$

Proposition: Given $\alpha, \beta : \mathbb{R}_+ \to \mathbb{R}_+$, $X, Y > 0$ s.t.:

$$\forall t > 0 : \quad \mathbb{P}\left(X \geq t\right) \leq \alpha(t) \quad \text{and} \quad \mathbb{P}\left(Y \geq t\right) \leq \beta(t)$$

Then $\mathbb{P}\left(X \cdot Y \geq t\right) \leq 2\alpha \boxtimes \beta(t)$

Proof: Denoting $\gamma \equiv \alpha \boxtimes \beta = (\alpha^{-1} \cdot \beta^{-1})^{-1}$, $\forall t > 0$:

$$\begin{aligned}
\mathbb{P}\left(X \cdot Y \geq t\right) &\leq \mathbb{P}\left(X \cdot Y \geq \alpha^{-1}(\gamma(t)) \cdot \beta^{-1}(\gamma(t))\right) \\
&\leq \mathbb{P}\left(X \geq \alpha^{-1}(\gamma(t))\right) + \mathbb{P}\left(Y \geq \beta^{-1}(\gamma(t))\right) \\
&\leq 2\gamma(t)
\end{aligned}$$

II - Operation with concentration inequalities.

Theorem: Consider $Z \in \mathbb{R}^n$, random, s.t. $\forall f : \mathbb{R}^n \to \mathbb{R}$, 1-Lipschitz:

$$\mathbb{P}\left(|f(Z) - f(Z')| \geq t\right) \leq \alpha(t) \qquad (Z, Z' \text{ i.i.d.})$$

- Consider $\Lambda : \mathbb{R}^n \to \mathbb{R}_+$ s.t.:

$$\forall t > 0 : \quad \mathbb{P}\left(\Lambda(Z) \geq t\right) \leq \beta(t)$$

- Consider $\Phi : \mathbb{R}^n \longrightarrow \mathbb{R}^p$ s.t. $\forall z, z' \in \mathbb{R}^n$:

$$\|\Phi(z) - \Phi(z')\| \leq \max(\Lambda(z), \Lambda(z'))\|z - z'\|$$

Then: $\forall f : \mathbb{R}^n \to \mathbb{R}$ 1-Lipschitz, $\forall t > 0$:

$$\forall t > 0 : \quad \mathbb{P}\left(|f(\Phi(Z)) - f(\Phi(Z')| \geq t\right) \leq 3\, \alpha \boxtimes \beta(t)$$

Proof: Denote $\Lambda = \Lambda(Z)$, $\Lambda' = \Lambda(Z')$, $\gamma \equiv \alpha \boxtimes \beta = (\alpha^{-1} \cdot \beta^{-1})^{-1}$, $\theta \equiv \beta^{-1}(\gamma(t))$

$$\mathbb{P}\left(|f(\Phi(Z)) - f(\Phi(Z')| \geq t\right) \leq \underbrace{\mathbb{P}\left(|f(\Phi(Z)) - f(\Phi(Z')| \geq t, \max(\Lambda, \Lambda') \leq \theta\right)}_{\leq \mathbb{P}(|h(\Phi(Z)) - h(\Phi(Z')| \geq t) \ \leq \alpha\left(\frac{t}{\beta^{-1}(\gamma(t))}\right)} + \underbrace{\mathbb{P}\left(\max(\Lambda, \Lambda') \geq \theta\right)}_{\leq 2\beta(\beta^{-1}(\gamma(t))}$$

$$\leq \alpha(\alpha^{-1}(\gamma(t)))$$

With $h : x \mapsto \sup_{\Lambda(z) \leq \theta} f \circ \Phi(z) - \theta d(x, z)$

$\to$ equal to $f \circ \phi$ on $\{z \in \mathbb{R}^n, \Lambda(z) \leq \theta\}$.

$\to$ $\beta^{-1}(\gamma(t))$-Lipschitz on $\mathbb{R}^n$

(Since $\forall t > 0 : \ \alpha^{-1}(\gamma(t)) \cdot \beta^{-1}(\gamma(t)) = t$)

II - Operation with concentration inequalities.

Theorem:

- Consider $Z \in \mathbb{R}^n$, s.t. $\forall f : \mathbb{R}^p \to \mathbb{R}$, 1-Lipschitz, convex:

$$\mathbb{P}\left(|f(Z) - f(Z')| \geq t\right) \leq \alpha(t)$$

with $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$.

Then, Given $d \in \mathbb{N}$, $\Phi : \mathbb{R}^n \longrightarrow \mathbb{R}^p$ d-times differentiable:

$$\mathbb{P}\left(\|\Phi(Z) - m_0\| \geq t\right) \leq C_d \; \alpha \circ \min_{k \in [d]} \left(\frac{c_d \; t}{m_k}\right)^{\frac{1}{k}},$$

where, $\forall k \in [d-1]$, we introduced m_k, a median of $\|d^k \Phi|_Z\|$ and $m_d \equiv \sup_{z \in \mathbb{R}^n} \|d^d \Phi|_z\|$.

Radosław **Adamczak** and Paweł **Wolff**. *Concentration inequalities for non-lipschitz functions with bounded derivatives of higher order.* Probability Theory and Related Fields, 162:531–586, 2015.

Friedrich **Götze**, Holger **Sambale**, and Arthur **Sinulis**. *Concentration inequalities for plynomials in sub-exponential random variables* Electron. J. Probab. 26: 1-22 (2021).

Application 1: Heavy tailed concentration

Proposition:

Consider $X = (X_1, \dots, X_n) \in \mathbb{R}^n$ and $Z = (Z_1, \dots, Z_n) \sim \mathcal{N}(0, I_n)$ such that:

- $\forall i \in [n] : \; X_i = \phi_i(Z_i)$ and $\forall t \in \mathbb{R} : \; |\phi_i'(t)| \leq \frac{C}{|t|} \exp(\frac{t^2}{2q})$

Then For all $f : \mathbb{R}^n \mapsto \mathbb{R}$ 1-Lipschitz:

$$\mathbb{P}\left(|f(X) - f(X')| \geq t\right) \leq \frac{C'n}{t^q}$$

Where C' only depends on C, q.

Application 1: Heavy tailed concentration - *Proof*

Theorem (Recall): Consider $Z \in \mathbb{R}^n$, random, s.t. $\forall f : \mathbb{R}^n \to \mathbb{R}$, 1-Lipschitz:

$$\mathbb{P}\left(|f(Z) - f(Z')| \geq t\right) \leq \alpha(t) \qquad (Z, Z' \text{ i.i.d.})$$

- Consider $\Lambda : \mathbb{R}^n \to \mathbb{R}_+$ s.t.:

$$\forall t > 0 : \quad \mathbb{P}\left(\Lambda(Z) \geq t\right) \leq \beta(t)$$

- Consider $\Phi : \mathbb{R}^n \longrightarrow \mathbb{R}^p$ s.t. $\forall z, z' \in \mathbb{R}^n$:

$$\|\Phi(z) - \Phi(z')\| \leq \max(\Lambda(z), \Lambda(z'))\|z - z'\|$$

Then: $\forall f : \mathbb{R}^n \to \mathbb{R}$ 1-Lipschitz, $\forall t > 0$:

$$\forall t > 0 : \quad \mathbb{P}\left(|f(\Phi(Z)) - f(\Phi(Z')| \geq t\right) \leq 3\, \alpha \boxtimes \beta(t)$$

Assume: $Z = (Z_1, \dots, Z_n) \sim \mathcal{N}(0, I_n) \leftrightarrow \ \alpha = \mathcal{E}_2 : t \mapsto 2e^{-t^2/2}$

$\forall i \in [n] : \ X_i = \phi_i(Z_i)$ and $\forall t \in \mathbb{R} : \ |\phi_i'(t)| \leq \frac{C}{|t|} \exp(\frac{t^2}{2q}) \ \equiv h(t)$

$$\|\Phi(Z) - \Phi(Z')\|^2 = \textstyle\sum_{i=1}^n |\phi_i(Z_i) - \phi_i(Z_i')|^2 \quad = \max_{1\leq i\leq n}(h(Z_i), h(Z_i'))^2 \sum_{i=1}^n |Z_i - Z_i'|^2$$

$$\mathbb{P}(\max_{1\leq i\leq n}(h(Z_i), h(Z_i')) \geq t) \leq 2n\mathbb{P}(Z_i \geq h^{-1}(t)) \leq 2n\mathcal{E}_2 \circ h^{-1}(t) \quad \equiv \beta(t)$$

$$\alpha \boxtimes \beta \leq (2n\mathcal{E}_2) \boxtimes 2n\mathcal{E}_2 \circ h^{-1} \leq 2n\mathcal{E}_2 \circ (\mathrm{Id} \boxtimes h)$$

$$\leq 2n\mathcal{E}_2 \circ (\mathrm{Id} \cdot h)^{-1} \leq 4n \exp\left(-\frac{1}{2}(\sqrt{2q \log(\mathrm{Id}/C)})^2\right) = \frac{4C^q n}{\mathrm{Id}^q}$$

Application 1: Heavy tailed concentration

Proposition:

Consider $X = (X_1, \dots, X_n) \in \mathbb{R}^n$ and $Z = (Z_1, \dots, Z_n) \sim \mathcal{N}(0, I_n)$ such that:

- $\forall i \in [n] : \; X_i = \phi_i(Z_i)$ and $\forall t \in \mathbb{R} : \; |\phi_i'(t)| \leq \frac{C}{|t|} \exp(\frac{t^2}{2q})$

Then For all $f : \mathbb{R}^n \mapsto \mathbb{R}$ 1-Lipschitz:

$$\mathbb{P}\left(|f(X) - f(X')| \geq t\right) \leq \frac{C' n}{t^q}$$

Where C' only depends on C, q.

NB: $\forall r < q : \mathbb{E}[|X_i|^r] < \infty$ and $\mathbb{E}[|f(X)|^r] < \infty$

Application 2: Hanson Wright Theorem

Theorem: (Hanson Wright) Given $A \in \mathcal{M}_n$ deterministic, $Z = (z_1, \ldots, z_n) \in \mathbb{R}^n$ such that:

- $\forall f : \mathbb{R}^n \to \mathbb{R}$ 1-Lipschitz:
  $\mathbb{P}\left(|f(Z) - \mathbb{E}[f(Z)]| \geq t\right) \leq \alpha(t/\eta)$
- With $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$, $\int t\alpha(t)dt < \infty$
- $\|\mathbb{E}[Z]\| \leq K$

$$\mathbb{P}\left(\left|Z^T AZ - \mathbb{E}[Z^T AZ]\right| \geq t\right) \leq C\alpha\left(-\frac{ct}{\eta\|A\|_F}\right) + C\alpha\left(\sqrt{\frac{ct}{\eta^2\|A\|}}\right)$$

Application 2: Hanson Wright Theorem *- Proof*

Theorem (Recall):

- Consider $Z \in \mathbb{R}^n$, s.t. $\forall f : \mathbb{R}^p \to \mathbb{R}$, 1-Lipschitz, convex:

$$\mathbb{P}\left(|f(Z) - \mathbb{E}[f(Z)]| \geq t\right) \leq \alpha(t)$$

with $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ and $\int t^d \alpha(t) dt \leq \infty$

Then, Given $d \in \mathbb{N}$, $\Phi : \mathbb{R}^n \longrightarrow \mathbb{R}^p$ d-times differentiable:

$$\mathbb{P}\left(|\Phi(Z) - \mathbb{E}[\Phi(Z)]| \geq t\right) \leq C_d \ \alpha \circ \min_{1 \leq k \leq d} \left(\frac{t}{m_k}\right)^{\frac{1}{k}},$$

where, $\forall k \in [d-1]$: $m_k = \mathbb{E}[\|d^k \Phi|_Z\|]$
and $m_d \equiv \sup_{z \in \mathbb{R}^n} \|d^d \Phi|_z\|$.

If $\int \alpha \leq \infty$:
$\mathbb{P}\left(|f(Z) - \mathbb{E}[f(Z)]| \geq t\right) \leq \alpha(t)$
$\implies \mathbb{P}\left(|f(Z) - f(Z')| \geq t\right) \leq C\alpha(ct)$
$\implies \mathbb{P}\left(|f(Z) - \mathbb{E}[f(Z)]| \geq t\right) \leq C'\alpha(c't$
For $C, C', c', c > 0$ numerical constant.

Theorem:

- Consider $Z \in \mathbb{R}^n$, s.t. $\forall f : \mathbb{R}^p \to \mathbb{R}$, 1-Lipschitz, convex:

$$\mathbb{P}\left(|f(Z) - \mathbb{E}[f(Z)]| \geq t\right) \leq \alpha(t)$$

with $\alpha : \mathbb{R}^+ \to \mathbb{R}_+$ and $\int t^d \alpha(t) dt \leq \infty$

Then, Given $d \in \mathbb{N}$, $\Phi : \mathbb{R}^n \longrightarrow \mathbb{R}^p$ d-times differentiable:

$$\mathbb{P}\left(|\Phi(Z) - \mathbb{E}[\Phi(Z)]| \geq t\right) \leq C_d \ \alpha \circ \min_{1 \leq k \leq d} \left(\frac{t}{m_k}\right)^{\frac{1}{k}},$$

where, $\forall k \in [d-1]$: $m_k = \mathbb{E}[\|d^k \Phi|_Z\|]$

and $m_d \equiv \sup_{z \in \mathbb{R}^n} \|d^d \Phi|_z\|$.

Set $\Phi : X \mapsto X^T A X$:

$$\forall H \in \mathbb{R}^n : d\phi|_X \cdot H = X^T A H + H^T A X \quad \text{and} \quad d^2\phi|_X \cdot H = 2H^T A H$$

$$\|d\phi|_X\| = 2\|AX\| \quad \text{and} \quad m_2 = \|d^2\phi|_X\| = \|A\|$$

$$m_1 = \mathbb{E}[\|d\phi|_X\|] = \mathbb{E}\|AX\| \leq \sqrt{\mathbb{E}[X^T A A^T X]} = \sqrt{\text{Tr}(\mathbb{E}[X X^T] A A^T)} \leq \|\mathbb{E}[X X^T]\| \ \|A\|_F$$

Application 2: Hanson Wright Theorem

Theorem: (Hanson Wright) Given $A \in \mathcal{M}_n$ deterministic, $Z = (z_1, \dots, z_n) \in \mathbb{R}^n$ such that:

- $\forall f : \mathbb{R}^n \to \mathbb{R}$ 1-Lipschitz:
  $$\mathbb{P}\left(|f(Z) - \mathbb{E}[f(Z)]| \geq t\right) \leq \alpha(t/\eta)$$
- With $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$, $\quad \int t\alpha(t)dt < \infty$
- $\|\mathbb{E}[Z]\| \leq K$

$$\mathbb{P}\left(\left|Z^T AZ - \mathbb{E}[Z^T AZ]\right| \geq t\right) \leq C\alpha\left(-\frac{ct}{\eta\|A\|_F}\right) + C\alpha\left(\sqrt{\frac{ct}{\eta^2\|A\|}}\right)$$

Application 3: Random matrix concentration

Given $x_1, \ldots, x_n$, independent random vectors, denote $X \equiv (x_1, \ldots, x_n) \in \mathbb{R}^{p \times n}$.

Goal: Eigen value distribution of $\frac{1}{n} X X^T$: $\mu \equiv \frac{1}{p} \sum_{i=1}^{p} \delta_{\lambda_i}$ **??**

Eigen values of $\frac{1}{p} X X^T$

$$\left(\text{Sp}\left(\frac{1}{p} X X^T \right) = \{\lambda_1, \ldots, \lambda_p\} \right)$$

- Correspondance $\mu \longleftrightarrow m : z \mapsto \int_{\mathbb{R}} \frac{1}{z - \lambda} d\mu(\lambda)$

"Steiltjes Transform" (similar to Cauchy Transform)

- Link with the *"Resolvent"*: $m(z) = \frac{1}{p} \mathrm{Tr} Q(z)$, where $Q(z) \equiv \left(z I_p - \frac{1}{n} X X^T \right)^{-1}$.

Strategy: Find deterministic $\tilde{Q} \in \mathcal{M}_p$ such that $Q \approx \tilde{Q}$

Application 3: Random matrix concentration

1- Concentration of $Q = (zI_n - \frac{1}{n}XX^T)^{-1}$

Assume $\forall f : \mathcal{M}_p \to \mathbb{R}$, 1-Lipschitz: $\mathbb{P}(|f(X) - f(X')| \geq t) \leq \alpha(t/\eta_{np})$

$M \mapsto (zI_p - \frac{1}{n}MM^T)^{-1}$ is $\frac{C}{\Im(z)\sqrt{n}}$-Lipschitz

$\implies \forall f : \mathcal{M}_p \to \mathbb{R}$, 1-Lipschitz:

$$\mathbb{P}\left(|f(Q) - \mathbb{E}[f(Q)]| \geq t\right) \leq \alpha(\sqrt{n}t/C\eta_{n,p})$$

Assume $\Im(z) \geq O(1)$

2- Find deterministic computable $\tilde{Q}$ close to $\mathbb{E}[Q]$.

Will deduce: $\forall A \in \mathcal{M}_p$ deterministic:

$$\mathbb{P}\left(|\mathrm{Tr}(A(Q - \tilde{Q}))| \geq t\right) \leq C\alpha(?)$$

Application 3: Random matrix concentration

Goal: Approach $\mathbb{E}[Q] = \mathbb{E}\left[\left(zI_p - \frac{1}{n}XX^T\right)^{-1}\right]$

- Of course $\mathbb{E}[Q]$ far from $(zI_p - \Sigma)^{-1}$ $\quad$ $\Sigma \equiv \frac{1}{n}\sum_{i=1}^n \Sigma_i$ where $\Sigma_i = \mathbb{E}\left[\frac{1}{n}x_i x_i^T\right]$, $\forall i \in [n]$

Solution: Look for $\tilde{Q} \equiv \left(zI_p - \Sigma^\Delta\right)^{-1}$ $\quad$ $\Sigma^\Delta \equiv \frac{1}{n}\sum_{i=1}^n \Delta_i \Sigma_i$, Δ to be determined

Given $A \in \mathcal{M}_p$, deterministic:

Dependence between Q and x_i

$$\mathrm{Tr}\left(A(\mathbb{E}[Q] - \tilde{Q})\right) = \mathbb{E}\left[\mathrm{Tr}\left(AQ\left(\Sigma^\Delta - \frac{1}{n}XX^T\right)\tilde{Q}\right)\right] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\mathrm{Tr}\left(\Delta_i AQ\Sigma_i\tilde{Q} - AQx_i x_i^T\tilde{Q}\right)\right]$$

Application 3: Random matrix concentration

$$\mathrm{Tr}\left(A(\mathbb{E}[Q]-\tilde{Q}_\delta)\right)=\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\mathrm{Tr}\left(\left(\Delta_i-\frac{1}{1+\frac{1}{n}x_i^TQ_{-i}x_i}\right)AQ_{-i}x_ix_i^T\tilde{Q}^\Delta\right)\right]+O\left(\frac{1}{\sqrt{n}}\right)$$

Use the *Schur Formula:* $Qx_i=\dfrac{Q_{-i}x_i}{1+\frac{1}{n}x_i^TQ_{-i}x_i}$, with $Q_{-i}\equiv\left(zI_p-\frac{1}{n}XX^T-x_ix_i^T\right)^{-1}$.

Independent with x_i

1. Chose $\Delta_i^{(1)}\equiv\mathbb{E}\left[\frac{1}{1-\frac{1}{n}x_i^TQ_{-i}x_i}\right]\approx\frac{1}{1-\frac{1}{n}\mathrm{Tr}(\Sigma\tilde{Q}^{\Delta^{(1)}})}$

   Relies on **Hanson-Wright Inequality:**

$$\mathbb{P}\left(\left|x_i^T\tilde{Q}^\Delta AQ_{-i}x_i-\mathbb{E}[x_i^T\tilde{Q}^\Delta AQ_{-i}x_i]\right|\geq t\right)\leq C\alpha\left(-\frac{ct}{\eta_p\|A\|_F}\right)+C\alpha\left(\sqrt{\frac{ct}{\eta_p^2\|A\|}}\right)$$

2. Chose $\Delta^{(2)}$ solution to $\Delta_i^{(2)}=\frac{1}{1-\frac{1}{n}\mathrm{Tr}(\Sigma\tilde{Q}^{\Delta^{(2)}})}$

Application 3: Random matrix concentration

Recall the objects: $X = (x_1, \ldots, x_n) \in \mathcal{M}_{p,n}$

$$Q = \left(zI_p - \frac{1}{n}XX^T\right)^{-1} \qquad \tilde{Q} \equiv \left(zI_p - \Sigma^\Delta\right)^{-1} \quad \Sigma^\Delta \equiv \frac{1}{n}\sum_{i=1}^n \Delta_i \Sigma_i$$

With Δ solution to $\Delta_i = \frac{1}{1 - \frac{1}{n}\mathrm{Tr}(\Sigma \tilde{Q}^\Delta)}$

Theorem: Assume $p \leq Cn$ and:

- $\forall f : \mathbb{R}^p \to \mathbb{R}$, 1-Lipschitz: $\mathbb{P}(|f(x_i) - f(x_i')| \geq t) \leq \alpha(t/\eta_p)$
- $\forall f : \mathcal{M}_{p,n} \to \mathbb{R}$, 1-Lipschitz: $\mathbb{P}(|f(X) - f(X')| \geq t) \leq \alpha(t/\sqrt{n}\eta_p)$
- $x_1, \ldots, x_n$ independents
- $\|\Sigma_i\| \leq C$

Then: if $\displaystyle\int t^3 \alpha(t)dt < \infty : \ \|\mathbb{E}[Q] - \tilde{Q}^\Delta\|_{HS} \leq C\frac{\eta_p}{\sqrt{n}}$

if $\displaystyle\int t\alpha(t)dt < \infty : \ \|\mathbb{E}[Q] - \tilde{Q}^\Delta\|_* \leq C\eta_p\sqrt{p}$

Application 3: Random matrix concentration

Proposition (Recall):

Consider $X = (X_1, \dots, X_n) \in \mathbb{R}^n$ and $Z = (Z_1, \dots, Z_n) \sim \mathcal{N}(0, I_n)$ such that:

- $\forall i \in [n] : \ X_i = \phi_i(Z_i)$ and $\forall t \in \mathbb{R} : \ |\phi_i'(t)| \leq \frac{C}{|t|} \exp(\frac{t^2}{2q})$

Then For all $f : \mathbb{R}^n \mapsto \mathbb{R}$ 1-Lipschitz:

$$\mathbb{P}\left(|f(X) - f(X')| \geq t\right) \leq \frac{C' n}{t^q} = \alpha(t/\eta_n) \text{ with } \alpha : t \mapsto \frac{C}{t^q} \text{ and } \eta_n = n^{1/q}$$

NB: $\forall r < q : \mathbb{E}[|X_i|^r] < \infty$ and $\mathbb{E}[|f(X)|^r] < \infty$

$$\int t^r \alpha(t) dt < \infty \iff q > r + 1 \iff \eta_n = o(n^{1/r+1}).$$

$$\int t \alpha(t) dt < \infty \iff q > 2 \iff \eta_n = o(\sqrt{n}).$$

Application 3: Random matrix concentration

Let us consider $\alpha : t \mapsto \frac{C}{t^q}$ $\longleftrightarrow$ $\eta_p = O(p^{1/q}) = O(n^{1/q})$.

Theorem: Assume $p \leq Cn$ and:

- $\forall f : \mathbb{R}^p \to \mathbb{R}$, 1-Lipschitz: $\mathbb{P}(|f(x_i) - f(x_i')| \geq t) \leq \alpha(t/\eta_p)$
- $\forall f : \mathcal{M}_{p,n} \to \mathbb{R}$, 1-Lipschitz: $\mathbb{P}(|f(X) - f(X')| \geq t) \leq \alpha(t/\sqrt{n}\eta_p)$
- $x_1, \ldots, x_n$ independents
- $\|\Sigma_i\| \leq C$

Then: if $q > 4$: $\|\mathbb{E}[Q] - \tilde{Q}^\Delta\|_{HS} \leq o\left(\frac{1}{n^{1/4}}\right)$

if $q > 2$: $\|\mathbb{E}[Q] - \tilde{Q}^\Delta\|_* \leq C\eta_p\sqrt{p}$

$\rightarrow$ Consequence for Stieltjes transform $m(z) = \frac{1}{p}\mathrm{Tr}(Q)$ in heavy tailed setting:

$$\mathbb{P}\left(\left|\frac{1}{p}\mathrm{Tr}(Q) - \frac{1}{p}\mathrm{Tr}(\tilde{Q}^\Delta)\right| \geq t\right) \leq \alpha\left(\frac{t}{o(1)}\right) \underset{t\to\infty}{\longrightarrow} 0$$