# IT for Statistics and Learning
## 2024

**Assignment 11**
Assigned: Friday, February 2, 2024
Due: Thursday, February 8, 2024 M. Skoglund

---

**Problem 11.1:** A random variable $X$ is *sub-exponential* with parameters $(\tau^2, b)$ if for all $\lambda$ such that $|\lambda| \leq 1/b$ we have
$$E[e^{\lambda(X-E[X])}] \leq e^{\lambda^2\tau^2/2}$$
Let $Z \sim \mathcal{N}(0,1)$ and set $X = Z^2$. Prove that $X$ is sub-exponential with $\tau^2 = b = 4$

**Problem 11.2:** Let $X \in \mathbb{R}^{m\times p}$ be genereted with iid entries $\sim \mathcal{N}(0, 1/m)$. Prove that for any $\varepsilon \in (0,1)$, $\delta \in (0,1)$,
$$m \geq \frac{8}{\varepsilon^2}\left[2\ln n + \ln\frac{1}{\delta}\right]$$
and any $n$ points $\{\theta_i\}$ in $\mathbb{R}^p$ we have that
$$(1-\varepsilon)\|\theta_i - \theta_j\|^2 \leq \|X\theta_i - X\theta_j\|^2 \leq (1+\varepsilon)\|\theta_i - \theta_j\|^2$$
with probability at least $1-\delta$, using the bound from Problem 11.1.

**Problem 11.3:** Consider the model $y = X\theta + z$ where $y$ has dimension $n$ and $\theta$ has dimension $p$, $p > n$, but $\|\theta\|_0 = k < p$. Let $g(y)$ denote any algorithm that produces $\hat{\theta} = g(y)$ from $y$. Fix $g(\cdot)$ and assume that there is a constant $C$ such that
$$\|g(X\theta + z) - \theta\|_2 \leq C\|z\|_2$$
for any $z \in \mathbb{R}^n$. Prove that then we have
$$\|b\|_2 \leq C\|Xb\|_2$$
for all $b \in \mathbb{R}^p$ for which $\|b\|_0 \leq 2k$

**Problem 11.4:** Consider now instead the model $y = X\theta^*$ (no noise and equality in $y = X\theta$ for $\theta = \theta^*$) where $y$ has dimension $n$ and $\theta$ has dimension $p$, $p > n$, but $\|\theta\|_0 = k < p$. Let $\theta_i^*$ be the elements of $\theta^*$ and let $J = \{i : |\theta_i^*| > 0\}$ be the set of indices for the support where $\theta_i^*$ is non-zero. Let $N(X)$ denote the null-space of the matrix $X$, i.e. $N(X) = \{a \in \mathbb{R}^p : Xa = 0\}$, and let $C(J) = \{a \in \mathbb{R}^d : \|a_{J^c}\|_1 \leq \|a_J\|_1\}$, i.e. the set of vectors for which the elements with indices in $J$ dominates the elements with indices in $J^c$. Consider the convex program
$$\min_{\theta} \|\theta\|_1 \quad \text{s.t.} \quad X\theta = y$$
and let $\hat{\theta}$ be its solution. Prove that $\hat{\theta} = \theta^*$ if $C(J) \cap N(X) = \{0\}$.

**Problem 11.5:** Consider a set $A \subset \mathbb{R}^d$. Let $\{Z_i\}_{i=1}^d$ be iid $\sim \mathcal{N}(0,1)$ and let
$$G(A) = E\left[\sup_{a\in A}\sum_{i=1}^d Z_i a_i\right]$$
denote the *Gaussian complexity* of the set $A$. Let
$$L_d(r) = \{\theta \in \mathbb{R}^d : \|\theta\|_0 \leq r \;\text{ and }\; \|\theta\|_2 \leq 1\}$$

Then it can be shown that

$$G(L_d(r)) \lesssim \sqrt{r \log \frac{ed}{r}}$$

Now, consider a matrix $X \in \mathbb{R}^{n \times d}$ such that

$$\gamma_1 \|h\|_2^2 \leq \frac{1}{n}\|Xh\|_2^2 \leq \gamma_2 \|h\|_2^2$$

for all $h \in \mathbb{R}^d$ s.t. $\|h\|_0 \leq 2r$ and assume that we observe $y = X\theta^* + z$ where $z$ has iid $\mathcal{N}(0,1)$ entries and where $\|\theta^*\|_0 = r$. Consider the program

$$\min_\theta \|y - X\theta\|_2^2 \;\text{ s.t. }\; \|\theta\|_0 = r$$

and let $\hat{\theta}$ be the (unique) solution. Prove that

$$\|\hat{\theta} - \theta^*\|_2^2 \lesssim \frac{\gamma_2}{\gamma_1^2} \frac{r \log(ed/r)}{n}$$

with a probability that approaches one as $r \log(ed/r) \to \infty$, using the bound on the Gaussian complexity of $L_d(r)$ (that you can use without proof).