1 Factorizations, Error sensitivity

1. If $\mathbf{Q} \times \mathbf{R} = \mathbf{X}$ is the QR décomposition of $\mathbf{X}$, check that $\mathbf{Q} \times \mathbf{Q}^T$ is the orthogonal projection on the subspace generated by the columns of $\mathbf{X}$.
2. Check the little perturbation bound for the OLS solution $\widehat{\theta}$ of $\arg\min_\theta \|\mathbf{X}\theta - y\|_2$ when y is perturbed by $\delta(y)$:
$$\|\delta(\widehat{\theta})\|_2 \leq \kappa(\mathbf{X}) \frac{\|\delta(y)\|_2}{\|\mathbf{X}\|_2} .$$
Hint : use SVD of $\mathbf{X}$.
3. Check the complete perturbation bound for the OLS solution $\widehat{\theta}$ of $\arg\min_\theta \|\mathbf{X}\theta - y\|_2$ when y is perturbed by $\delta(y)$:
$$\|\delta(\widehat{\theta})\|_2 \leq \kappa(\mathbf{X}) \frac{\|\delta(y)\|_2}{\|\mathbf{X}\|_2} .$$
Hint : use SVD of $\mathbf{X}$.
4. Solving least square problems (minimize $\|\mathbb{X}\theta - Y\|_2^2$) by computing $\left(\mathbb{X}^T\mathbb{X}\right)^{-1}\mathbb{X}^T$ may lead to numerical problems. Take
$$\mathbb{X} = \begin{pmatrix} 1 & 1 \\ \epsilon & 0 \\ 0 & \epsilon \end{pmatrix} .$$
Assume that ϵ is larger than machine precision while ϵ^2 is not
   (a) Compute $\mathbb{X}^T\mathbb{X}$. Is the machine able to invert $\mathbb{X}^T\mathbb{X}$?
   (b) What are the singular values of $\mathbb{X}$, of $\mathbb{X}^T\mathbb{X}$?
   (c) What is the condition number of $\mathbb{X}$, of $\mathbb{X}^T\mathbb{X}$?
   (d) What is the QR decomposition of $\mathbb{X}$? Does this QR decomposition suffer from the same precision problem as $\mathbb{X}^T\mathbb{X}$?

2 Stochastic inequalities

1. Maximum spacing in a uniform sample. Provide an upper bound on the expected length of the maximum spacing in an n-sample of the uniform distribution over $[0, 1]$. A spacing is the length of the interval between two adjacent sample points.
2. Kolmogorov-Smirnov statistics. The KS-statistics is defined as
$$Z_n := \sqrt{n} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|$$
where F_n is the empirical distribution function defined by a sample of n points collected independently from a probability distribution definef by distribution function F.
Establish an upper bound on $\mathbb{E}Z_n$, prove that there exists a universal constant C such that for all F and n, $\mathbb{E}Z_n \leq C$.
3. McDiarmid inequality (bounded-differences inequality).

Let $(\mathcal{X}, \mathcal{G}, P)$ be a probability space and $\Omega = \mathcal{X}^n, \mathcal{F} = \mathcal{G}^{\otimes n}, \mathbb{P} = P^{\otimes n}$, be the associated product space. Let f be a function from $\mathcal{X}^n \to \mathbb{R}$ such that there exists a sequence of constants $c_1, \ldots, c_n$ satisfying

$$\left| f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n) \right| \leq c_i \mathbb{I}_{x_i \neq x_i'}$$

for all $x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n, x_i'$.
Let $X_1, \ldots, X_n$ denote the random variables that map $\omega = (x_1, \ldots, x_n) \in \Omega = \mathcal{X}^n$, on $X_i(\omega) = x_i$.
Let $Z = f(X_1, \ldots, X_n)$.
Let $\mathcal{F}_i = \sigma(X_1, \ldots, X_i)$.
Let also $M_i = \mathbb{E}[Z \mid \mathcal{F}_i]$ for i from 0 to n.

(a) Why is $(\mathcal{F}_i)_{i \leq n}$ a filtration ?

(b) Show that $(M_i)_{i \leq n}$ is an $(\mathcal{F}_i)_{i \leq n}$-adapted martingale.

(c) Show that $\mathrm{var}[Z] \leq v := \sum_{i=1}^n c_i^2/4$.

(d) Show that for $t \geq 0$

$$\mathbb{P}\left\{Z - \mathbb{E}Z \geq t\right\} \leq \mathrm{e}^{-\frac{t^2}{2v}} .$$

4. Using McDiarmid inequality. Let $\mathcal{F}$ be a (finite) class of functions from $\mathcal{X}$ to $[0, 1]$. Let $X_1, \ldots, X_n$ be i.i.d. according to some probability distribution on $\mathcal{X}$. Let

$$Z_n := \sup_{f \in \mathcal{F}} \sum_{i=1}^n \frac{1}{n} \left(f(X_i) - \mathbb{E}f(X_i)\right)$$

(Z_n is a supremum of a bounded centered empirical process).
Check that Z_n satisfies the conditions of the bounded-differences inequality and that

$$\mathbb{P}\left\{Z_n - \mathbb{E}Z_n \geq t\right\} \leq \exp\left(-2nt^2\right) .$$