Spring School,
Westerland, Germany

Lectures by Martin Wainwright (wainwrig@berkeley.edu)
Based on book in preparation

Problem Set 1
March 2015

Problem 1.1

A zero-mean variable Z is sub-Gaussian with parameter σ if $\mathbb{E}[e^{\lambda Z}] \leq e^{\lambda^2 \sigma^2/2}$ for all $\lambda \in \mathbb{R}$. Let $\{Z_i\}_{i=1}^n$ be a sequence of zero-mean random variables, each sub-Gaussian with parameter σ. (No independence assumptions are needed.)

(a) Prove that $\mathbb{E}\big[\max_{i=1,\ldots,n} Z_i\big] \leq \sqrt{2\sigma^2 \log n}$ for all $n \geq 1$. (*Hint:* The exponential is a convex function.)

(b) Prove that $\mathbb{E}\big[\max_{i=1,\ldots,n} |Z_i|\big] \leq 2\sqrt{\sigma^2 \log n}$ for all $n \geq 2$.

Problem 1.2

For a given $q \in (0, 1]$, recall the (strong) ℓ_q-ball

$$\mathbb{B}_q(R_q) := \big\{\theta \in \mathbb{R}^d \mid \sum_{j=1}^d |\theta_j|^q \leq R_q\big\}. \tag{1}$$

The weak ℓ_q-ball with parameters (C, α) is defined as

$$\mathbb{B}_{w(\alpha)}(C) := \big\{\theta \in \mathbb{R}^d \mid |\theta|_{(j)} \leq C j^{-\alpha} \quad \text{for } j = 1, \ldots, d\big\}. \tag{2}$$

Here $|\theta|_{(j)}$ denote the order statistics of θ^* in absolute value, ordered from largest to smallest (so that $|\theta|_{(1)} = \max_{j=1,2,\ldots,d} |\theta_j|$ and $|\theta|_{(d)} = \min_{j=1,2,\ldots,d} |\theta_j|$.)

(a) Show that the set $\mathbb{B}_q(R_q)$ is star-shaped around the origin. (A set $\mathcal{C} \subseteq \mathbb{R}^d$ is star-shaped around the origin if $\theta \in \mathcal{C} \Rightarrow t\theta \in \mathcal{C}$ for all $t \in [0, 1]$.)

(b) For any $\alpha > 1/q$, show that there is a radius R_q depending on (C, α) such that $\mathbb{B}_{w(\alpha)}(C) \subseteq \mathbb{B}_q(R_q)$. This inclusion underlies the terminology "strong" and "weak" respectively.

(c) For a given integer $s \in \{1, 2, \ldots, d\}$, the best s-term approximation to a vector $\theta^* \in \mathbb{R}^d$ is given by

$$\Pi_s(\theta^*) := \arg \min_{\|\theta\|_0 \leq s} \|\theta - \theta^*\|_2^2. \tag{3}$$

Give a closed form expression for $\Pi_s(\theta^*)$.

(d) When $\theta^* \in \mathbb{B}_q(R_q)$ for some $q \in (0, 1]$, show that the best s-term approximation satisfies

$$\|\Pi_s(\theta^*) - \theta^*\|_2^2 \leq \big(R_q\big)^{2/q} \, \big(\frac{1}{s}\big)^{\frac{2}{q}-1}. \tag{4}$$

Problem 1.3

For a given design matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$, suppose that its columns $\{X_1, \ldots, X_d\}$ satisfy the normalization condition $\|X_j\|_2/\sqrt{n} = 1$ for all $j = 1, \ldots, d$. Define its pairwise incoherence $\mu(\mathbf{X}) := \max_{i \neq j} |\langle X_i, \, X_j \rangle / n|$. In this exercise, we prove that for a given sparsity s, the condition

$$s \, \mu(\mathbf{X}) < \gamma \qquad \text{for a sufficiently small constant } \gamma \tag{5}$$

implies that the restricted nullspace property holds.

(a) Let $S \subset \{1, 2, \ldots d\}$ be any subset of size s. Prove that the condition (5) implies there is a function $\gamma \mapsto c(\gamma)$ such that $\lambda_{\min}(\mathbf{X}_S^T \mathbf{X}_S / n) \geq c(\gamma) > 0$, as long as γ is sufficiently small.

(b) Prove that $\mathbf{X}$ satisfies the restricted nullspace property with respect to S. (Do this from first principles, without using any results on restricted isometry.)

Problem 1.4

Let $\mathbf{X} \in \mathbb{R}^{n \times d}$ be a standard Gaussian random matrix (i.e., $X_{ij} \sim N(0, 1)$, i.i.d. for all entries $(i.j)$).

(a) Letting $X_j \in \mathbb{R}^n$ be its j^{th} row, prove that there are constants c_1, c_2 such that

$$\mathbb{P}\Big[\max_{j \neq k} \big|\frac{\langle X_j, \, X_k \rangle}{n}\big| \geq \delta\Big] \leq c_1 e^{-c_2 n \delta^2} \qquad \text{for all } \delta \in (0, 1).$$

(*Hint:* The random variable $\frac{1}{\sqrt{n}}\big(\|X_j\|_2 - \mathbb{E}[\|X_j\|_2]\big)$ is sub-Gaussian with parameter $\sigma = 1/\sqrt{n}$.)

(b) Use this result to show that X satisfies the restricted nullspace property for all sparsity $s \leq c_3 \sqrt{\frac{n}{\log d}}$.

Problem 1.5

Consider the standard linear regression model $y = \mathbf{X}\theta^* + w$, where $\theta^* \in \mathbb{B}_q(R_q)$. Using the oracle inequality from lecture, and given an appropriate lower bound on the sample size n in terms of (d, R_q, σ, q), show that there are universal constants (c_0, c_1, c_2) such that with probability $1 - c_1 e^{-c_2 \log d}$, any Lasso solution $\widehat{\theta}$ satisfies the bound

$$\|\widehat{\theta} - \theta^*\|_2^2 \leq c_0 \, R_q \big(\frac{\sigma^2 \log d}{n}\big)^{1-\frac{q}{2}}.$$

Problem 1.6

Consider the sparse linear regression model $y = \mathbf{X}\theta^* + w$, where $w \sim \mathcal{N}(0, \sigma^2 I_{n\times n})$ and $\theta^* \in \mathbb{R}^d$ is supported on a subset S. Suppose that the sample covariance matrix $\widehat{\Sigma} = \frac{1}{n}\mathbf{X}^T\mathbf{X}$ has its diagonal entries uniformly upper bounded by one, and that for some parameter $\gamma > 0$, it also satisfies an ℓ_∞-curvature condition of the form

$$\big\|\widehat{\Sigma}\Delta\big\|_\infty \geq \gamma \|\Delta\|_\infty \quad \text{for all } \Delta \in \mathbb{C}_3(S). \tag{6}$$

Show that with the regularization parameter $\lambda_n = 4\sigma\sqrt{\frac{\log d}{n}}$, any Lasso solution satisfies the ℓ_∞-bound

$$\|\widehat{\theta} - \theta^*\|_\infty \leq \frac{6\sigma}{\gamma}\sqrt{\frac{\log d}{n}}$$

with high probability.

Problem 1.7

For an integer $k \in \{1, \ldots, d\}$, consider the following two subsets:

$$\begin{aligned}
\mathbb{L}_0(k) &:= \mathbb{B}_2(1) \cap \mathbb{B}_0(k) \;=\; \big\{\theta \in \mathbb{R}^d \mid \|\theta\|_2 \leq 1, \text{ and } \|\theta\|_0 \leq k\big\}, \\
\mathbb{L}_1(k) &:= \mathbb{B}_2(1) \cap \mathbb{B}_1(\sqrt{k}) \;=\; \big\{\theta \in \mathbb{R}^d \mid \|\theta\|_2 \leq 1, \text{ and } \|\theta\|_1 \leq \sqrt{k}\big\}.
\end{aligned}$$

Let $\overline{\text{conv}}$ denote the closure of the convex hull (when applied to a set).

(a) Prove that $\overline{\text{conv}}\big(\mathbb{L}_0(k)\big) \subseteq \mathbb{L}_1(k)$.

(b) Prove that $\mathbb{L}_1(k) \subseteq 3\,\overline{\text{conv}}\big(\mathbb{L}_0(k)\big)$.

(*Hint:* For part (b), you may find it useful to consider the support functions of the two sets.)

Problem 1.8

Let $\mathbf{X} \in \mathbb{R}^{n\times d}$ be a fixed design matrix such that $\frac{|\!|\!|\mathbf{X}_S|\!|\!|_{\text{op}}}{\sqrt{n}} \leq C$ for all subsets S of cardinality at most s. In this exercise, we show that with high probability, any solution of the constrained Lasso

$$\widehat{\theta} \in \arg\min_{\|\theta\|_1 \leq R} \Big\{\frac{1}{2n}\|y - \mathbf{X}\theta\|_2^2\Big\}$$

with $R = \|\theta^*\|_1$ satisfies the bound

$$\|\widehat{\theta} - \theta^*\|_2 \precsim \frac{\sigma}{\kappa}\sqrt{\frac{s\log(e\,d/s)}{n}} \tag{7}$$

where $s = \|\theta^*\|_0$. Note that this bound provides an improvement for linear sparsity (i.e., whenever $s = \alpha d$ for some constant $\alpha \in (0,1)$).

(a) Define the random variable

$$Z := \sup_{\Delta\in\mathbb{R}^d} \big|\langle \Delta, \frac{1}{n}\mathbf{X}^T w\rangle\big| \quad \text{such that } \|\Delta\|_2 \leq 1 \text{ and } \|\Delta\|_1 \leq \sqrt{s}, \tag{8}$$

where $w \sim \mathcal{N}(0, \sigma^2 I)$. Show that

$$\mathbb{P}\Big[Z \geq c_1\, C\,\sigma\,\Big\{\sqrt{\frac{s\log\frac{ed}{s}}{n}} + \delta\Big\}\Big] \leq c_2 e^{-c_3\, n\delta^2}$$

for universal constants (c_1, c_2, c_3). (*Hint:* The result of Exercise 1.7(b) may be useful here.)

(b) Use part (a) and results from lecture to show that if $\mathbf{X}$ satisfies an RE condition, then any optimal Lasso solution $\widehat{\theta}$ satisfies the bound (7) with probability $1 - c_2' e^{-c_3' s\log\left(\frac{ed}{s}\right)}$.