E9 203: Homework - 6

Assigned on: 19 Apr. 2015, due 30 Apr. 2015; answer any 10 questions.

1 Topics

- Subgaussian random variables
- Concentration of measure
- JL Lemma
- LASSO, uniqueness
- Strong convexity, implication to unconstrained convex optimization
- Compressible signals

Notation: "Triple-bar" norms $|||\cdot|||$ denote (vector norm) induced matrix norms while "double-bar" norms $\|\cdot\|$ denote vector norms (possibly, on matrices). For example, $|||\mathbf{A}|||_2$ is the same as $\|\mathbf{A}\|_{2\to 2}$ we used in class. Also, $\|\mathbf{A}\|_2$ or $\|\mathbf{A}\|_F$ is the matrix Frobenius norm.

2 Problems

1. Subgaussian moments. A random variable X is said to be subgaussian distributed if $\exists$ a constant $c > 0$ s.t. $\forall\, s > 0, \mathbb{E}\{e^{sX}\} \leqslant e^{cs^2}$. Show that $\exists$ a universal constant K s.t. if X is subgaussian, for every $q \in \mathbb{Z}_+$,
$$(\mathbb{E}\{|X|^q\})^{\frac{1}{q}} \leqslant K\sqrt{cq}.$$

2. Subgaussian moments – converse. Let X be a zero-mean random variable s.t. $\exists\, c > 0$ satisfying
$$(\mathbb{E}\{|X|^q\})^{\frac{1}{q}} \leqslant \sqrt{cq}$$
for every positive integer q. Show that X is subgaussian, i.e., for any $s > 0$,
$$\mathbb{E}\{e^{sX}\} \leqslant \sqrt{2}e^{1/6}e^{ces^2/2}$$

3. Let $X_1, \ldots, X_n$ be independent random variables, taking values from $[0, 1]$. Let $m \triangleq \mathbb{E}S_n$, where $S_n \triangleq \sum_{i=1}^n X_i$. Show that, for any $t \geqslant m$,
$$\mathbb{P}\{S_n \geqslant t\} \leqslant \left(\frac{m}{t}\right)^t \left(\frac{n-m}{n-t}\right)^{n-t}.$$
Hint. Use Chernoff's bounding method.

4. Use the previous result to show that

$$\mathbb{P}\{S_n \geqslant t\} \leqslant \left(\frac{m}{t}\right)^t e^{t-m},$$

and for all $\epsilon > 0$,

$$\mathbb{P}\{S_n \geqslant m(1+\epsilon)\} \leqslant e^{-mh(\epsilon)},$$

where $h(u) \triangleq (1+u)\log(1+u) - u$ for $u > 0$. Finally,

$$\mathbb{P}\{S_n \leqslant m(1-\epsilon)\} \leqslant e^{-m\epsilon^2/2}.$$

5. (J-L Lemma) Given $\epsilon > 0$ and $N \in \mathbb{Z}_+$. Let $m \in \mathbb{Z}_+$ be such that $m \geqslant O\left(\frac{\log N}{\epsilon^2}\right)$. For every set, $\mathcal{P}$ of N points in $\mathbb{R}^n$, there exists a mapping, $\psi : \mathbb{R}^n \to \mathbb{R}^m$, such that $\forall\ u, v \in \mathcal{P}$:

$$(1-\epsilon)\|u-v\|_2^2 \leqslant \|\psi(u) - \psi(v)\|_2^2 \leqslant (1+\epsilon)\|u-v\|_2^2$$

6. (Completing a small step in the proof that Gaussian matrices satisfying RIP) Show that

$$\frac{\sqrt{1+\delta/\sqrt{2}}}{1-\delta/14} \leqslant \sqrt{1+\delta}$$

and

$$\sqrt{1-\delta/\sqrt{2}} - \delta\sqrt{1+\delta}/14 \geqslant \sqrt{1-\delta}$$

for $\delta > 0$.

7. (Also a small step in the proof that Gaussian matrices satisfy RIP) Let $\mathbf{A} \in \mathbb{R}^{m\times n}$, with $A_{ij} \sim \mathcal{N}(0, \frac{1}{m})$ and i.i.d. Show that

$$\Pr\left\{\|\mathbf{A}\mathbf{x}\|_2^2 \leqslant (1-\epsilon)\|\mathbf{x}\|_2^2\right\} \leqslant \exp\left(-\frac{m\epsilon^2}{12}\right)$$

8. (Concentration of measure inequality for strictly sub-Gaussian matrices) Recall that we say X is ssub(c^2) if X is zero mean, $\mathbb{E}(X^2) = c^2$, and $\mathbb{E}\left\{e^{sX}\right\} \leqslant e^{c^2s^2/2}\,\forall\, s \in \mathbb{R}$. Suppose $\mathbf{A} \in \mathbb{R}^{m\times n}$, with A_{ij} being i.i.d. and ssub$(\frac{1}{m})$. Let $Y = \mathbf{A}\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^n$. Then, for any $\epsilon > 0$, show that

$$\mathbb{E}\left\{\|Y\|_1^2\right\} = \|\mathbf{x}\|_2^2$$

$$\Pr\left\{\left|\|Y\|_2^2 - \|\mathbf{x}\|_2^2\right| \geqslant \epsilon\|\mathbf{x}\|_2^2\right\} \leqslant 2\exp\left(-\frac{m\epsilon^2}{k^*}\right),$$

with $k^* = \frac{2}{1-\log(2)}$.

9. First, using results we showed in class, establish the following theorem:

Theorem 1 *Let s be such that $\delta_{3s} + 3\delta_{4s} < 2$. Then, for any $\mathbf{x}_0$ with $\mathcal{T}_0 \triangleq supp(\mathbf{x}_0)$, and $|\mathcal{T}_0| \leqslant s$, and any perturbation $\mathbf{e}$ with $\|\mathbf{e}\| \leqslant \epsilon$, the solution $\mathbf{x}^*$ to*

$$\mathcal{P}_1^\epsilon : \min_{\mathbf{x}} \|\mathbf{x}\|_1 \ s.\ t.\ \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leqslant \epsilon$$

obeys

$$\|\mathbf{x}^* - \mathbf{x}_0\|_2 \leqslant C_s\epsilon$$

where the constant C_s can be chosen such that it depends only on δ_{4s}.

(a) Extend the above theorem to the case where $\mathbf{e}$ is Gaussian distributed with mean 0 and variance σ^2. You will now have a probability of failure: the procedure will fail whenever $\|\mathbf{e}\| > \epsilon$. Comment on the trade-off between the number of measurements m and the noise variance σ^2.

(b) Extend the theorem to the case where $\mathbf{A}$ has i.i.d. Gaussian entries with mean 0 and variance $1/m$. Again, you will now have a probability of failure associated with the procedure.

(c) What if $\mathbf{A}$ has i.i.d. $\mathcal{N}(0, 1/m)$ entries *and* $\mathbf{e}$ is i.i.d. $\mathcal{N}(0, \sigma^2)$ distributed? Comment on your result.

(d) (Bonus question) Find the above results in the literature.

10. Consider the observation model: $\mathbf{y} = \mathbf{X}\beta^* + \mathbf{w}$, where where $\mathbf{X} \in \mathbb{R}^{m\times n}$, $\mathbf{y} \in \mathbb{R}^m$ and $m \ll n$. $\mathbf{w} \in \mathbb{R}^m$ is the zero mean additive observation noise. Consider the unconstrained form of LASSO estimator:

$$\min_{\beta\in\mathbb{R}^n} \left\{ \frac{1}{2m}\|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda_m\|\beta\|_1 \right\} \tag{1}$$

(a) Show that there exists at least one solution of (1). Why is this solution not unique?
(Hint: Instead of unconstrained form, use the equivalent constrained forms.)

(b) (Uniqueness of LASSO solution.) Prove the following:

i. A vector $\hat{\beta} \in \mathbb{R}^n$ is an optimal solution for (1) iff there exists a sub-gradient vector, $\hat{\mathbf{z}} \in \partial\|\hat{\beta}\|_1$ such that

$$\frac{1}{m}\mathbf{X}^T\mathbf{X}(\hat{\beta} - \beta^*) - \frac{1}{m}\mathbf{X}^T\mathbf{w} + \lambda_m\hat{\mathbf{z}} = 0. \tag{2}$$

ii. Suppose that $\hat{\mathbf{z}}$ satisfies the strict dual feasibility condition $|\hat{\mathbf{z}}_j| < 1$ for all $j \notin S(\hat{\beta})$, where $S(.)$ denotes the support function. Then, show that any optimal solution $\tilde{\beta}$ to the LASSO satisfies $\tilde{\beta}_j = 0$ for all $j \notin S(\hat{\beta})$.
(Hint: Use the KKT conditions for optimality of a convex program. Also use the definition of sub-differential of ℓ_1-norm function.)

iii. Under the conditions of part(b), if the $k \times k$ matrix $\mathbf{X}_{S(\hat{\beta})}^T\mathbf{X}_{S(\hat{\beta})}$ is invertible, then $\hat{\beta}$ is the unique optimal solution of (1).

11. (Implications of strong convexity on unconstrained convex optimization problems.) Consider an unconstrained convex optimization problem:

$$\min_{\mathbf{x}} f(\mathbf{x}) \tag{3}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function. We further assume that $f(\mathbf{x})$ is strongly convex, i.e., $\exists\ m > 0$ such that $\nabla^2 f(\mathbf{x}) \geq m\mathbf{I}$. (Here, $\mathbf{A} \geq \mathbf{B}$ implies $\mathbf{A} - \mathbf{B}$ is positive semi-definite.) Prove

(a) For any $\mathbf{x}, \mathbf{y}$ in dom(f) we have:

$$f(\mathbf{y}) \geqslant f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) + \frac{m}{2}\|\mathbf{y} - \mathbf{x}\|_2^2. \tag{4}$$

(b) Use the above inequality to show:

$$f(\mathbf{x}) - f(\mathbf{x}^*) \leqslant \frac{1}{2m}\|\nabla f(\mathbf{x})\|_2^2. \tag{5}$$

where $\mathbf{x}^*$ is the optimal solution.
(Hint: In the above equation, choose even a smaller RHS by using a value of $\mathbf{y}$ that minimizes the RHS. Note that the RHS is quadratic in $\mathbf{y}$ with $\mathbf{x}$ fixed.)

(c) Further, prove that

$$\|\mathbf{x} - \mathbf{x}^*\|_2 \leqslant \frac{2}{m}\|\nabla f(\mathbf{x})\|_2. \tag{6}$$

(Hint: Use the first equation with $\mathbf{y} = \mathbf{x}^*$ and then use Cauchy-Schwarz inequality.)

Note: this means that under strong convexity conditions, closeness of the objective function to the optimal value implies the closeness of the solution vector to the optimal solution. For LASSO program, in the under-determined setting, this leads us to impose restricted strong convexity and restricted eigenvalue properties on the measurement matrix.

12. Consider the *overdetermined* LASSO, i.e., $m > n$ and also consider the case when we have orthonormal $\mathbf{X}$. Show that the solution for:

$$\min_{\beta \in \mathbb{R}^n} \left\{ \frac{1}{2}\|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda\|\beta\|_1 \right\} \tag{7}$$

is given by:

$$\beta_j^{\text{lasso}} = \text{sign}(\hat{\beta}_j^{LS})(|\hat{\beta}_j^{LS}| - \lambda)^+ \tag{8}$$

where $\hat{\beta}^{LS}$ is the least squares solution for $\mathbf{y} = \mathbf{X}\beta$.

13. Show that, for every signal $\mathbf{y} \in \mathbb{R}^n$ and every positive integer t, we have

$$\|\mathbf{y} - \mathbf{y}_t\|_2 \leqslant \frac{1}{2\sqrt{t}}\|\mathbf{y}\|_1, \tag{9}$$

where $\mathbf{y}_t$ is the signal in $\mathbb{R}^n$ that is formed by restricting $\mathbf{y}$ to its t largest-magnitude entries.
(Hint: A. Gilbert, M. Strauss, J. Tropp, and R. Vershynin, "One sketch for all: Fast algorithms for compressed sensing", Proc. 39th ACM Symp. Theory of Computing, Jun. 2007.)

14. A signal $\mathbf{x} \in \mathbb{R}^N$ is said to be p-compressible with magnitude R if the sorted components of the signal decay at the rate

$$|x|_{(i)} \leqslant R \cdot i^{-1/p}, \text{ for } i = 1, 2, 3, \dots \tag{10}$$

When $p = 1$, this implies $\|\mathbf{x}\|_1 \leqslant R(1 + \log N)$. When $p \approx 0$, this implies that $\mathbf{x}$ is very nearly sparse. Show that compressible signals can be approximated by sparse signals as follows:

$$\begin{aligned} \|\mathbf{x} - \mathbf{x}_s\|_1 &\leqslant \frac{p}{1-p} R s^{1-1/p} \\ \|\mathbf{x} - \mathbf{x}_s\|_2 &\leqslant \sqrt{\frac{p}{2-p}} R s^{1/2-1/p}. \end{aligned}$$

(Hint: Write each norm as a sum and then approximate the sum with an integral.)