EE127 - EE227AT 5/11/2017
L. El Ghaoui

Final: Solutions

1. (*10 points*) Let $X \in \mathbf{R}^{n \times m}$, $y \in \mathbf{R}^m$ and $\lambda, \mu > 0$. Consider the so-called (square-root) "elastic net" problem

$$\min_w \|X^T w - y\|_2 + \lambda \|w\|_1 + \mu \|w\|_2. \tag{1}$$

   (a) Let $k << \min(m, n)$ be given. Assume that we know the top k singular values σ_i, $i = 1, \ldots, k$, of X, and the associated left and right singular vectors, denoted $u_i \in \mathbf{R}^n$, $v_i \in \mathbf{R}^m$, $i = 1, \ldots, k$. Construct matrices $L \in \mathbf{R}^{n \times k}$, $R \in \mathbf{R}^{m \times k}$ such that $X \approx LR^T$, with $R^T R = I_k$ (the $k \times k$ identity matrix).

   (b) Replace X by its low-rank approximation in problem (1). Show that the new problem reduces to one with n variables and $k+1$ measurements (that is, the new vector y has dimension $k+1$). *Hint:* square and expand the first term in (1), and prove the fact that $I - RR^T$ is positive semi-definite.

   (c) In practice, it may be advisable to take into account the size of the error made in approximating X. How would you modify the problem to account for this error? *Hint:* you may assume the singular value σ_{k+1} is known, and use the fact that, for any $n \times m$ matrix Z, m-vector y and scalar $\epsilon > 0$:

$$\max_{\Delta \,:\, \|\Delta\| \le \epsilon} \|(Z + \Delta)^T w - y\|_2 = \|Z^T w - y\|_2 + \epsilon \|w\|_2,$$

   with $\|\Delta\|$ the largest singular value of matrix Δ.

   (d) The computational complexity of problem (1) grows in $O(nm^2 + m^3)$. That of forming the k first singular values for an $n \times m$ matrix grows as $O(nmk)$. What is the complexity estimate of the problem you obtained in part 1c? Make sure to account for the overhead involved in forming the reduced problem. (If you are not sure about part 1c, use the formulation you found in part 1b.)

Solution:

(a) We have

$$\begin{aligned}
\|X^T w - y\|_2^2 &= \|RL^T w - y\|_2^2 \\
&= w^T L R^T R L^T w - 2 y^T R L^T w + \|y\|_2^2 \\
&= \|L^T w - R^T y\|_2^2 + s^2 = \left\| \begin{pmatrix} L^T \\ 0 \end{pmatrix} w - \begin{pmatrix} R^T y \\ s \end{pmatrix} \right\|_2^2,
\end{aligned}$$

where $s^2 := y^T (I - RR^T) y$. Note that the latter quantity is indeed non-negative, since the fact $R^T R = I_k$ guarantees that the eigenvalues of RR^T are either 1 or 0, hence $RR^T \preceq I_m$.

It thus suffices to solve the reduced square-root LASSO problem:

$$\min_w \|\tilde{L}^T w - \tilde{y}\|_2 + \lambda \|w\|_1 + \mu \|w\|_2,$$

where

$$\tilde{L} = \begin{pmatrix} L & 0 \end{pmatrix}, \;\; \tilde{y} = \begin{pmatrix} R^T y \\ s \end{pmatrix},$$

which has n variables and $k + 1$ measurements.

(b) We can solve the robust version

$$\min_w \max_{\Delta \,:\, \|\Delta\| \le \epsilon} \|(Z+\Delta)^T w - y\|_2 + \lambda \|w\|_1 + \mu \|w\|_2.$$

We have $\|X - LR^T\| \le \sigma_{k+1} := \epsilon$, hence

$$\max_{\Delta \,:\, \|\Delta\| \le \epsilon} \|(Z+\Delta)^T w - y\|_2 = \|Z^T w - y\|_2 + \epsilon \|w\|_2,$$

We are led to solve the reduced-size "elastic net" problem

$$\min_w \|L^T w - R^T y\|_2 + \lambda \|w\|_1 + (\mu + \epsilon)\|w\|_2.$$

(c) The computational complexity of the reduced problem, including the cost of the partial SVD and that of forming $\tilde{y}$, grows as $O(mnk + nk^2 + k^3)$. This represents substantial computational savings, for example when $n = m = 10^4$, $k = 500$, the reduced problem can be solved twenty times faster.

2. (*10 points*) The log-sump-exp function is defined as the function $f : \mathbf{R}^n \to \mathbf{R}$ with values

$$f(z) = \log\left(\sum_{i=1}^{n} e^{z_i}\right).$$

Often, the convexity of this function is proven by deriving the Hessian. In this exercise, you will show that f is convex, using another method.

(a) Show that, for any given $s > 0$, we have

$$\log s = -1 + \min_v \; se^v - v.$$

(b) Show that

$$f(z) = -1 + \min_v \sum_{i=1}^{n} e^{z_i+v} - v.$$

(c) Prove convexity of f based on the above result. *Hint:* use the notion of joint convexity.

Solution:

(a) Since $s > 0$, the differentiable function $v \to se^v - v$ is convex. The optimality condition of the problem of minimizing this function is that its gradient is zero. Differentiating, we obtain that $v^* = -\log s$ is optimal, hence the result.

(b) This follows from applying the result of the previous part to $s = \sum_{i=1}^{n} e^{z_i}$.

(c) The function

$$(z, v) \in \mathbf{R}^n \times \mathbf{R} \to \sum_{i=1}^{n} e^{z_i+v} - v$$

is jointly convex in its arguments z, v; hence the minimum over v is convex in z.

3. (*10 points*) *Fermat point on a triangle:* Consider a triangle with distinct vertices A_1, A_2, A_3 in two dimensions. We seek the point P that minimizes the sum of the distances to the vertices of the triangle. Representing the points as three distinct vectors $a_i \in \mathbf{R}^2$, $i = 1, 2, 3$, and the (unknown) location of P as $x \in \mathbf{R}^2$, the problem writes as

$$p^* = \min_x \sum_{i=1}^{3} \|x - a_i\|_2.$$

In this exercise, we show that the point P satisfies $\widehat{A_1PA_2} = \widehat{A_2PA_3} = \widehat{A_3PA_1} = 120°$.

(a) Use Sion's theorem to show that the problem admits the dual expression:

$$p^* = \max_{y_1,y_2,y_3} \sum_{i=1}^{3} a_i^T y_i \ : \ y_1 + y_2 + y_3 = 0, \quad \|y_i\|_2 \leq 1, \quad i = 1, 2, 3.$$

(b) Assume that the primal optimal point x^* is unique, and that $x^* \neq a_i$ for every $i = 1, 2, 3$. Show that optimal dual variables y_i^*, $i = 1, 2, 3$ satisfy

$$y_i^* = \frac{a_i - x^*}{\|x^* - a_i\|_2}, \quad i = 1, 2, 3.$$

Hint: show that strong duality implies that

$$\sum_{i=1}^{3} (a_i - x^*)^T y_i^* = \sum_{i=1}^{3} \|x^* - a_i\|_2.$$

and reason by contradiction, invoking Cauchy-Schwartz.

(c) Show that the dual solution vectors satisfy $y_i^T y_j = -1/2$, $i \neq j$. *Hint:* Express $\|y_1\|_2^2$ as a function of y_2, y_3.

(d) Conclude the argument.

Solution:

(a) We have

$$p^* = \min_x \max_{(y_1,y_2,y_3)\in\mathcal{Y}} \sum_{i=1}^{3} y_i^T (a_i - x).$$

where

$$\mathcal{Y} = \left\{ (y_1, y_2, y_3) \in \mathbf{R}^{2\times 3} \ : \ \|y_i\|_2 \leq 1, \quad i = 1, 2, 3 \right\}.$$

Since the objective function in the min-max problem above is convex-concave, and $\mathcal{Y}$ is convex and compact, we can invoke Sion's theorem, which leads to

$$p^* = \max_{(y_1,y_2,y_3)\in\mathcal{Y}} \min_x \sum_{i=1}^{3} y_i^T (a_i - x).$$

This yields the desired dual expression.

(b) At optimum, we have, for any optimal dual points y_i^*, $i = 1, 2, 3$:

$$\sum_{i=1}^{3} a_i^T y_i^* = \sum_{i=1}^{3} \|x^* - a_i\|_2.$$

Using the fact that the vectors y_i^* sum to zero:

$$\sum_{i=1}^{3} (a_i - x^*)^T y_i^* = \sum_{i=1}^{3} \|x^* - a_i\|_2. \tag{2}$$

Using the Cauchy-Schwartz inequality, we have

$$(a_i - x^*)^T y_i^* \leq \|x^* - a_i\|_2 \|y_i\|_2 \leq \|x^* - a_i\|_2, \quad i = 1, 2, 3.$$

Assume that one of the above inequalities is strict; summing, we obtain a contradiction with (2). We conclude that the inequalities above are all equalities, which in turn proves that

$$y_i^* = \frac{a_i - x^*}{\|a_i - x^*\|_2}, \quad i = 1, 2, 3. \tag{3}$$

(c) We have

$$1 = \|y_1\|_2^2 = \|y_2 + y_3\|_2^2 = \|y_2\|_2^2 + \|y_3\|_2^2 - 2y_2^T y_3 = 2(1 - y_2^T y_3).$$

This shows that $y_2^T y_3 = -1/2$, so that the angle between the two unit-norm vectors y_2, y_3 is indeed $120°$. We prove the value of the other angles similarly.

(d) According to (3), the angle between the vectors y_1 and y_2 is the same as the angle between the vectors $a_1 - x^*$ and $a_2 - x^*$, which is the angle between PA and PB.

4. (*10 points*) We consider an investment problem of the form

$$p^* = \max_{x \geq 0} \; r^T x - \frac{1}{2} x^T C x$$

where $r \in \mathbf{R}^n$ is a vector of expected returns, $C = C^T \succ 0$ is a covariance matrix. We assume that C is given as a so-called "single factor model", that is

$$C = D + ff^T,$$

where D is diagonal positive-definite, and $f \in \mathbf{R}^n$.

(a) First show a preliminary result: for every scalars ρ and $\delta > 0$,

$$\max_{\xi \geq 0} \; \rho\xi - \frac{1}{2}\delta\xi^2 = \frac{1}{2\delta} \max(0, \rho)^2.$$

with unique minimizer $\xi^* = \max(0, \rho)/\delta$.

(b) Based on the expression

$$p^* = \max_{x \geq 0} \; r^T x - \frac{1}{2} x^T D x - \frac{1}{2} z^2 \; : \; z = f^T x,$$

show that a dual can be written as

$$\frac{1}{2} \min_{\nu} \; \nu^2 + \sum_{i=1}^{n} \frac{\max(0, r_i - \nu f_i)^2}{D_{ii}}.$$

Hint: for the minimization of the Lagrangian over x, use part 4a .

(c) Does strong duality hold? Justify your answer.

(d) Explain how to recover an optimal primal dual point from an optimal dual variable ν^*.

Solution:

(a) The unconstrained minimizer is unique, and given by

$$\xi^* = \frac{\rho}{\delta}.$$

If this point is feasible (that is, $\rho \geq 0$) then it is optimal; if not, $\xi^* = 0$ is optimal. Hence the maximizer is unique, and given by

$$\xi^* = \frac{\max(0, \rho)}{\delta}.$$

The desired expression follows.

(b) We have

$$p^* = \max_{x \geq 0} r^T x - \frac{1}{2} x^T D x - \frac{1}{2} z^2 + \nu(z - f^T x).$$

A dual can be obtained by exchanging the min and max in the above:

$$p^* \leq d^* := \min_{\nu} \max_{x \geq 0} r^T x - \frac{1}{2} x^T D x - \frac{1}{2} z^2 + \nu(z - f^T x).$$

Maximizing over z leads to the unique solution (for any fixed ν) $z^*(\nu) = \nu$. Maximizing over x leads to a problem of the form seen in part 4a, with $\rho = (r_i - \nu f_i)$, $\delta = D_{ii}$ (here the index i is given). Hence

$$\min_{x_i \geq 0} (r_i - \nu f_i) x_i - \frac{1}{2} D_{ii} x_i^2 = \frac{\max(0, r_i - \nu f_i)^2}{D_{ii}}, \quad i = 1, \ldots, n,$$

with unique minimizer

$$x_i^*(\nu) = \frac{\max(0, r_i - \nu f_i)}{D_{ii}}, \quad i = 1, \ldots, n. \tag{4}$$

Plugging in the optimal points $x^*(\nu)$ and $z^*(\nu)$ in the Lagrangian leads to the desired dual.

(c) Strong duality is a direct consequence of Slater's theorem, since the primal is convex, has equality constraints only and is feasible.

(d) We use the fact that for every ν, the minimizer of the Lagrangian over x is unique. This implies that an optimal point x^* can be set as in (4), with ν set to be the optimal dual point.