APM462H1S, Winter 2014 , Assignment 2,

due: Monday February 24, at the beginning of the lecture.

Exercise 1. Assume that Q is a symmetric $n \times n$ matrix, with eigenvalues $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_n$ and with an *orthonormal basis* of eigenvectors $w_1, \ldots, w_n$.

Since $w_1, \ldots, w_n$ is a basis, any vector $v \in E^n$ can be written in the form

$$v = a_1 w_1 + \cdots + a_n w_n. \tag{1}$$

(In fact, $a_i = w_i^T v$ for every i — this follows by multiplying equation (1) by w_i^T on the left and using the fact that the vectors $w_1, \ldots, w_n$ are orthonormal.)

a. Show that if $v = a_1 w_1 + \cdots + a_n w_n$ and at least one a_i is nonzero, then

$$\frac{v^T Q v}{v^T v} = \theta_1 \lambda_1 + \ldots + \theta_n \lambda_n, \qquad \text{where } \theta_i = \frac{a_i^2}{a_1^2 + \cdots + a_n^2}.$$

solution sketch: Just write

$$v^T Q v = (a_1 w_1 + \cdots + a_n w_n)^T Q (a_1 w_1 + \cdots + a_n w_n) = \sum_{i,j=1}^n a_i a_j w_i^T Q w_j,$$

$$v^T v = (a_1 w_1 + \cdots + a_n w_n)^T (a_1 w_1 + \cdots + a_n w_n) = \sum_{i,j=1}^n a_i a_j w_i^T w_j$$

and simplify, using the facts that

$$Q w_i = \lambda_i w_i, \qquad w_i^T w_j = \begin{cases} 1 & \text{if } i = j \\ & \text{if not.} \end{cases}$$

b. Using part **a** (if you like), prove that

$$\lambda_n = \text{largest eigenvalue of } Q = \max_{v \ne 0} \frac{v^T Q v}{v^T v}. \tag{2}$$

solution. Since $\lambda_i \le \lambda_n$ for all i, and since $\theta_i \ge 0$ for all i, for every nonzero $v \in E^n$ we have

$$\frac{v^T Q v}{v^T v} = \theta_1 \lambda_1 + \ldots + \theta_n \lambda_n \le \theta_1 \lambda_n + \ldots + \theta_n \lambda_n = (\theta_1 + \ldots + \theta_n) \lambda_n = \lambda_n.$$

Thus

$$\max_{v \ne 0} \frac{v^T Q v}{v^T v} \le \lambda_n$$

On the other hand,

$$\frac{w_n^T Q w_n}{w_n^T w_n} = \lambda_n$$

so that

$$\max_{v \ne 0} \frac{v^T Q v}{v^T v} \ge \lambda_n.$$

remark. By almost the same argument, one can also show that

$$\lambda_1 = \text{smallest eigenvalue of } Q = \min_{v \ne 0} \frac{v^T Q v}{v^T v}. \tag{3}$$

Exercise 2. Assume that Q is a symmetric $n \times n$ matrix, $c \in E^n$ is a nonzero (column) vector, and μ is a positive number.

Consider the symmetric matrix $R = Q + \mu cc^T$.

Let $\lambda_i(Q)$ denote the ith eigenvalue of Q, and similarly and $\lambda_i(R)$ the ith eigenvalue of R, where they are arranged so that $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$, for both Q and R.

a. Prove that

$$\lambda_n(R) \geq \mu|c|^2 + \lambda_1(Q).$$

solution: By formula (2)

$$\lambda_n(R) = \max_{v \neq 0} \frac{v^T R v}{v^T v} \geq \frac{c^T R c}{c^T c} = \frac{c^T Q c}{c^T c} + \mu \frac{c^T c c^T c}{c^T c}$$

By formula (3), $\frac{c^T Q c}{c^T c} \geq \lambda_1(Q)$, and since $c^T c = |c|^2$, we deduce from the above that

$$\lambda_n(R) \geq \lambda_1(Q) + \mu \frac{(|c|^2)^2}{|c|^2} = \lambda_1(Q) + \mu |c|^2.$$

b. Prove that if $n \geq 2$, then

$$\lambda_1(R) \leq \lambda_n(Q).$$

solution. If $n \geq 2$, then there must be a nonzero vector $w \in E^n$ such that $w^T c = 0$. For this vector, $w^T R w = w^T Q w$. Thus by formulas (3) and (2) (in that order),

$$\lambda_1(R) \leq \frac{w^T R w}{w^T w} = \frac{w^T Q w}{w^T w} \leq \lambda_n(Q).$$

c. Conclude that if Q is positive semidefinite, then the condition number of R satisfies

$$\text{condition number of } R = \frac{\lambda_n(R)}{\lambda_1(R)} \geq \frac{\mu|c|^2}{\lambda_n(Q)}.$$

Thus, the condition number is very large if μ is large compared to $\lambda_n(Q)$.

solution. If Q is positive semidefinite, then $\lambda_1(Q) \geq 0$, and part **a** implies that $\lambda_n(R) \geq \mu|c|^2$. So it immediately follows that

$$\text{condition number of } R = \frac{\lambda_n(R)}{\lambda_1(R)} \geq \frac{\mu|c|^2}{\lambda_n(Q)}.$$

Exercise 3. Luenberger and Ye, problem 21 on page 260. In the definition of f in the book, s^2 should be replaced by x^2.

a, b. Find an unconstrained local minimum point of $f(x, y) = x^2 + xy + y^2 - 3x$ and explain why it is a global minimum point.

solution. The first-order conditions are:

$$\frac{\partial f}{\partial x} = 2x + y - 3 = 0, \qquad \frac{\partial f}{\partial y} = 2y + x = 0.$$

The only solution of this system of equations is

$$\boxed{(x^*, y^*) = (2, -1)}$$

So this is the only critical point of f. Also, the Hessian matrix of f is

$$\nabla^2 f = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

which is easily seen[1] to be positive definite. Thus every critical point is in fact a global minimum point. In particular, $(2, -1)$ is a global minimum.

c. Find the minimum point of f subject to $x \geq 0, y \geq 0$.
solution. There are four cases:
case 1. the minimum occurs where $x > 0$ and $y > 0$.
This is impossible, since if this were the case, then the first-order conditions would be $\nabla f = 0$; but we already know that the only point where $\nabla f = 0$ does not satisfy $x > 0, y > 0$.

case 2. the minimum occurs at $(0, 0)$.
Then the first-order conditions are:

$$\frac{\partial f}{\partial x} \geq 0, \qquad \frac{\partial f}{\partial y} \geq 0,$$

But $\frac{\partial f}{\partial x} = -3$ at $(0, 0)$ so these are not satisfied and this point is not a local minimum.
case 3 the minimum occurs where $x = 0, y > 0$ Then the first-order conditions are:

$$\frac{\partial f}{\partial x} = 0 \qquad \frac{\partial f}{\partial y} \geq 0,$$

If $x = 0$ then $\frac{\partial f}{\partial y} = 2y$, so the point where $x = 0$ and $\frac{\partial f}{\partial y} = 0$ is the point $(0, 0)$, which we already know is not a local minimum.
case 4 the minimum occurs where $x > 0, y = 0$ Then the first-order conditions are:

$$\frac{\partial f}{\partial x} \geq 0 \qquad \frac{\partial f}{\partial y} = 0,$$

If $y = 0$ then $\frac{\partial f}{\partial x} = 2x - 3$, so the only possible minimum point is

$$\boxed{(x^*, y^*) = (\frac{3}{2}, 0)}$$

and this in fact is the global minimum.

Exercise 4. Luenberger and Ye, problem 24 on page 260, parts **a - c**.
Extra marks will be awarded for a correct solution of part **d**, which is harder.
solution:
a. The first-order conditions are: if a local minimum point of f is attained at $x^* = (x_1^*, \ldots, x_n^*)$, then

$$\frac{\partial f}{\partial x_i}(x^*) = 0 \text{ if } x_i^* > 0, \qquad \frac{\partial f}{\partial x_i}(x^*) \geq 0 \text{ if } x_i^* = 0,$$

b. Suppose that $d = (d_1, \ldots, d_n) = 0$ at some point.
Then if you just stare at the algorithm as it appears in the textbook, you can see that the necessary conditions are satisfied.

---

[1]In fact, for every matrix of the form $\begin{pmatrix} a & b \\ b & a \end{pmatrix}$, the eigenvalues are $a + b$ and $a - b$, with eigenvectors $\binom{1}{1}$ and $\binom{1}{-1}$ respectively.

In particular, if $d_i = 0$ then, since it cannot be the case that both $-g_i = 0$ and $g_i < 0$, it must be the case that for every i,

either $x_i > 0$ and $-g_i = -\frac{\partial f}{\partial x_i}(x^*) = 0$ ***or*** $x_i = 0$ and $g_i = \frac{\partial f}{\partial x_i}(x^*) \geq 0$.

This is equivalent to the necessary conditions.

c. Suppose that $d \neq 0$ at x.

Then by a first-order Taylor approximation,

$$f(x - sd) = f(x) - s\nabla f(x)d + o(s|d|),$$

using the "little oh" notation. And by the definition of d,

$$\nabla f(x)d = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} d_i = \sum_{i=1}^{n} d_i^2 = |d|^2 > 0.$$

So

$$f(x - sd) - f(x) = -s|d|^2 + o(s|d|),$$

and the right-hand side is negative for all sufficiently small values of s. This shows that there exist some $s > 0$ such that $f(x - sd) < f(x)$.

d. The Global Convergence Theorem does not apply to this algorithm, becuase it is not closed.

To see this suppose that $n = 2$, and that $f(x) = f(x_1, x_2) = x_1 + x_2$.

Then the following is true:

if $x_1 \geq x_2 > 0$ **then** $A(x_1, x_2) = \{(x_1 - x_2, 0)$**.**$\}$

This is true because in this case, $d = (1, 1)$, so $f(x - sd) = f(x_1 - s, x_2 - s) = x_1 + x_2 - 2s$.

We have to minimize this over all choice of s for which $x - sd$ satisfies the constraints, or equivalently, both $x_1 - s \geq 0$ and $x_2 - s \geq 0$.

Clearly, the function $f(x - sd)$ is minimized by choosing the largest value of s consistent with the constraints, and this choice is $s = x_2$. So the minimum occurs at $(x_1 - x_2, 0)$.

if $x_2 \geq x_1 > 0$ **then** $A(x_1, x_2) = \{(0, x_2 - x_1)\}$**.**

The reasoning is similar to the previous case.

if $x_1 = 0$ **or** $x_2 = 0$**, then** $A(x_1, x_2) = \{(0, 0)\}$**.**

Let us suppose for concreteness that $x_1 = 0$ and $x_2 > 0$. Then $d = (0, 1)$, so $x - ds = (0, x_2 - s)$ and $f(x - sd) = x_2 - s$. As before, this is minimized by choosing s to be as large as possible, consistent with the constraints, and this choice is $s = x_2$. So the minimum occurs at $(0, 0)$. The other cases are similar.

Now we can easily see that the algorithm is not closed. For example, let $x^k = (1, \frac{1}{k})$ and $y^k = (1 - \frac{1}{k}, 0)$. Then $x^k \to x = (1, 0)$ and $y^k \to y = (1, 0)$, but $y \notin A(x)$, since $A(x) = \{(0, 0\}$.