CO 367 Fall 2018: Homework 1

Due: October 5th, 1:30pm at the start of the lecture

Instructions For every nontrivial step you perform, you must justify why the step is valid and what assumption it exploits. In other words, you do not need to justify basic algebraic operations (rearranging or distributing terms, multiplying both sides of an equation by a constant, etc.), but you *do* need to explain all steps that exploit hypotheses and assumptions (positive semidefiniteness of a matrix, continuity or convexity of a function, taking a limit that must exist, etc.).

Question 1 Let $A \in \mathbb{R}^{n\times n}$ be a symmetric, positive semidefinite matrix, and let k be a positive integer.

a. [4 marks] Construct a matrix $G \in \mathbb{R}^{n\times n}$ such that $G^k = A$ for any given A and k. Prove that $G^k = A$.

b. [2 marks] Show that G is invertible if and only if A is invertible.

c. [2 marks] By giving an example, show that G is not necessarily unique in satisfying $G^k = A$.

Solution: a. Assume that the eigenvalues of A are $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ and that the corresponding eigenvectors are given by $x_1, \ldots, x_n$. Since A is symmetric, there exists an orthogonal matrix $Q = [x_1|\cdots|x_n]$ and a diagonal matrix $D = \text{diag}(\lambda_1, \ldots, \lambda_n)$ such that $A = QDQ^T$. We construct $G := QD^{1/k}Q^T$ where $D^{1/k} = \text{diag}(\lambda_1^{1/k}, \ldots, \lambda_n^{1/k})$. This is possible because with real numbers, since A is positive semidefinite, $\lambda_1, \ldots, \lambda_n \geq 0$. Then, $G^n = (QD^{1/k}Q^T)(QD^{1/k}Q^T)\cdots(QD^{1/k}Q^T)$. Since $Q^TQ = I$, we have $G^n = QD^{1/k}\cdots D^{1/k}Q^T = QDQ^T = A$.

Note: An alternative construction for a. uses $A = \sum_i \lambda_i x_i^T x$. Similarily, we construct $G = \sum_i \lambda_i^{1/k} x_i^T x$.

b. Since $G = QD^{1/k}Q^T$ and Q is invertible, G is invertible if and only if $D^{1/k}$ is invertible. This happens if and only if $\lambda_i \neq 0$ for all i, which is the case if and only if A is invertible.

c. For simplicity, we will use $k = 2$. Let us construct a positive semidefinite matrix A as follows. First, we construct an orthogonal matrix $Q = \begin{bmatrix} 3/5 & 4/5 \\ -4/5 & 3/5 \end{bmatrix}$. The columns of Q will be eigenvectors of A. Then we choose eigenvalues, say 100 and 25. We have $A = QDQ^T = \begin{bmatrix} 52 & -36 \\ -36 & 73 \end{bmatrix}$ and $G = QD^{1/2}Q^T = \begin{bmatrix} 34/5 & -12/5 \\ -12/5 & 41 \end{bmatrix}$. We can then verify that not only $GG = A$, but also $(-G)(-G) = A$.

Question 2 [6 marks] Prove the following. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^2-smooth function over $\mathbb{R}^n$, and let $\alpha \in \mathbb{R}$ be some constant. If $(\nabla f(y) - \nabla f(x))^T(y - x) \geq \alpha||y - x||_2^2$ for all $x, y \in \mathbb{R}^n$, then $(\nabla^2 f(z) - \alpha I)$ is positive semidefinite for all $z \in \mathbb{R}^n$.

Solution: Take any $t \in \mathbb{R}$ and $w \in \mathbb{R}^n$. Using our hypothesis with $y = x + tw$, we get that

$$(\nabla f(x + tw) - \nabla f(x))^T tw \geq \alpha t^2||w||^2.$$

Let $g : \mathbb{R}^n \to \mathbb{R}$ be given by the ith component of $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$, i.e. $g(x) = (\nabla f(x))_i$. Since $f \in C^2(\mathbb{R}^n)$, we know that $g \in C^1(\mathbb{R}^n)$. We can thus write, using Taylor's 1st-order expansion, that

$$g(x+tw) = g(x) + tw^T \nabla g(x) + \phi(tw) = g(x) + \nabla g(x)^T tw + \phi(tw),$$

where $\lim_{t\to 0} \frac{\phi(tw)}{t} = 0$. Writing the above equation for all i, we get

$$\nabla f(x+tw) = \nabla f(x) + \nabla^2 f(x)^T tw + \psi(tw),$$

where $\psi : \mathbb{R}^n \to \mathbb{R}^n$ satisfies $\lim_{t\to 0} \frac{\psi(tw)}{t} = \mathbf{0}$. Coming back to our hypothesis, we get

$$(\nabla f(x) + \nabla^2 f(x)^T tw + \psi(tw) - \nabla f(x))^T tw \geq \alpha t^2 ||w||^2,$$

thus,

$$(\nabla^2 f(x)^T tw)^T tw + \psi(tw)^T tw \geq \alpha t^2 ||w||^2.$$

Dividing by t^2, using $||w||_2^2 = w^T w$, we get

$$w^T \nabla^2 f(x) w + \frac{\psi(tw)^T}{t} w \geq \alpha w^T w.$$

Taking the limit for $t \to 0$,

$$w^T \nabla^2 f(x) w \geq \alpha w^T w,$$

then,

$$w^T \nabla^2 f(x) w - \alpha w^T w \geq 0,$$

and finally

$$w^T (\nabla^2 f(x) - \alpha I) w \geq 0.$$

Since this is still true for any $x, w \in \mathbb{R}^n$, $\nabla^2 f(x) - \alpha I$ is positive semidefinite, for all $x \in \mathbb{R}^n$.

Alternative solution: Compute the Taylor expansion of $f(x+h)$ at x and the Taylor expansion of $f(x)$ at $x+h$. We get

$$f(x+h) = f(x) + h^T \nabla f(x) + \frac{1}{2} h^T \nabla^2 f(x+\lambda h) h \tag{1}$$

$$f(x) = f(x+h) - h^T \nabla f(x+h) + \frac{1}{2} h^T \nabla^2 f(x+\sigma h) h \tag{2}$$

for some $0 \leq \lambda, \sigma \leq 1$. Add (1) and (2) together. We get

$$\begin{aligned} \frac{1}{2} h^T (\nabla^2 f(x+\lambda h) + \nabla^2 f(x+\sigma h)) h &= h^T (\nabla f(x+h) - \nabla f(x)) \\ &\geq \alpha ||h||_2^2 \end{aligned}$$

We can now set $h = tw$ where $t \geq 0$ and $||w||_2 = 1$, and proceed as in the other proof.

Question 3 Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be two convex functions over their respective domains. Moreover, assume that g is monotonically nondecreasing.

a. [3 marks] Show that if we define $h : \mathbb{R}^n \to \mathbb{R}$ as

$$h(x) = g(f(x)),$$

then h is convex over $\mathbb{R}^n$.

b. [3 marks] Let $Q \in \mathbb{R}^{n\times n}$ be a symmetric, positive semidefinite matrix, and let $\beta > 0$ be a positive scalar. Use a. to show that $q : \mathbb{R}^n \to \mathbb{R}$ defined by

$$q(x) = e^{\beta x^T Qx}$$

is convex over $\mathbb{R}^n$. (If you use the fact that some function is convex, you must prove that fact.)

Solution: a. We need to show that g is convex, i.e., for all $x, y \in \mathbb{R}^n$ and for all $0 \le \alpha \le 1$,

$$h(\alpha x + (1-\alpha)y) \le \alpha h(x) + (1-\alpha)h(y),$$

or, replacing $h(x)$ by its expression,

$$g(f(\alpha x + (1-\alpha)y)) \le \alpha g(f(x)) + (1-\alpha)g(f(y)).$$

Since g is nondecreasing, we have for all $a, b \in \mathbb{R}$,

$$a \le b \quad \Rightarrow \quad g(a) \le g(b). \tag{3}$$

Since f is convex, we have, for all $x, y \in \mathbb{R}^n$ and for all $0 \le \alpha \le 1$,

$$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha)f(y), \tag{4}$$

Using (4) as the left-hand side of the implication (3), we get

$$g(f(\alpha x + (1-\alpha)y)) \le g(\alpha f(x) + (1-\alpha)f(y)). \tag{5}$$

The convexity of g gives

$$g(\alpha a + (1-\alpha)b) \le \alpha g(a) + (1-\alpha)g(b). \tag{6}$$

Letting $a := f(x)$ and $b := f(y)$ in (6), we get an upper bound on the right-hand side of (5):

$$g(\alpha f(x) + (1-\alpha)f(y)) \le \alpha g(f(x)) + (1-\alpha)g(f(y)),$$

so finally

$$g(f(\alpha x + (1-\alpha)y)) \le \alpha g(f(x)) + (1-\alpha)g(f(y)).$$

b. We use a. with $g(x) := e^x$ and $f(x) := \beta x^T Qx$. Now we only need to show that (i) e^x is monotonically increasing, (ii) e^x is convex, and (iii) $\beta x^T Qx$ is convex.

(i) $\frac{\partial}{\partial x} e^x = e^x \ge 0$ so e^x is monotonically increasing.

(ii) $\nabla^2 e^x = \frac{\partial^2}{\partial x^2} e^x = e^x > 0$ for all $x \in \mathbb{R}$. In this case the Hessian is a 1×1 matrix that is positive semidefinite, because $y e^x y = y^2 e^x \geq 0$ for all $y \in \mathbb{R}$, so e^x is convex.

(iii) First we establish an expression for $\nabla^2 \beta x^T Q x = \frac{\partial^2}{\partial x^2} \beta x^T Q x$. We have

$$\beta x^T Q x = \sum_i x_i \sum_j Q_{ij} x_j = \sum_{ij} Q_{ij} x_i x_j = \sum_i Q_{ii} x_i^2 + 2 \sum_{i<j} Q_{ij} x_i x_j,$$

since Q is symmetric. We get

$$\frac{\partial^2}{\partial x^2} x^T Q x = 2 Q_{ij},$$

so $\nabla^2 \beta x^T Q x = 2\beta Q$, which is a constant positive semidefinite matrix, by hypothesis. Therefore, $\beta x^T Q x$ is convex.