Modern methods of decision making 2021
Seminar : February 8

Exercise 1

Consider a matrix $A \in \mathbb{R}^{d\times d}$ and the function $f : \mathbb{R}^d \to \mathbb{R}$ defined by $f(x) = x^T A x$.
Under which assumptions is f convex, or α - convex ?

Solution

f isn't convex for any choice of A.
Indeed in the simple case where $d = 1$, $A = a$ is simply a real number and $f(x) = ax^2$ which is convex only if $a \geqslant 0$.
More generally, $\forall\ d \geqslant 1$, the function f is infinitely many times differentiable and its Hessian matrix is, $\forall x \in \mathbb{R}^d$:

$$\nabla^2 f(x) := A + A^T$$

So f is α - convex, for $\alpha \geqslant 0$, iff :

$$\forall\ h \in \mathbb{R}^d : \quad h^T (A + A^T)\, h \geqslant \alpha \|h\|^2.$$

In the case where A is symmetric ($A = A^T$), this condition translates as

$$\forall\ h \in \mathbb{R}^d : \ h^T A\, h \geqslant \frac{\alpha}{2} \|h\|^2.$$

Let us prove that

$$\forall h \in \mathbb{R}^d : h^T A h \geqslant \frac{\alpha}{2} \|h\|^2 \quad \overset{(*)}{\Longleftrightarrow} \quad \text{all eigenvalues of } A \text{ are } \geqslant \frac{\alpha}{2}.$$

First, we recall the Spectral Theorem :

Thm (Spectral Theorem)

Let $A \in \mathbb{R}^{d \times d}$ be a symmetric matrix.
Then all eigenvalues of A belong to $\mathbb{R}$ and can be ordered as $\lambda_1(A) \geqslant \lambda_2(A) \geqslant \dots \geqslant \lambda_d(A)$.
In addition, there exists an orthormal basis $\epsilon_1, \dots, \epsilon_d$ of $\mathbb{R}^d$ that diagonalizes A, i.e., such that :

$$\forall\, 1 \leqslant i \leqslant d, \quad A\epsilon_i = \lambda_i(A)\, \epsilon_i$$

Remark : The last statement can be written equivalently as follows : there exists an invertible matrix P, such that $P^{-1} = P^T$ and such that

$$A = P \operatorname{diag}(\lambda_1(A), \dots, \lambda_d(A))\, P^{-1}$$

Back to the equivalence $(*)$

$\Rightarrow$) This implication is rather straightforward. Indeed, taking any eigenvalue λ of A and any associated eigenvector $h \in \mathbb{R}^d \setminus \{0\}$ ($Ah = \lambda h$), we have

$$\lambda \|h\|^2 = h^T A h \underset{\substack{\uparrow \\ \text{assumption}}}{\geqslant} \frac{\alpha}{2} \|h\|^2 \implies \lambda \geqslant \frac{\alpha}{2}.$$

$\Leftarrow$) Conversely suppose $\lambda_i(A) \geq \frac{\alpha}{2}$, $\forall i = 1, \dots, n$.
Then let $\epsilon_1, \dots, \epsilon_d$ be as in the spectral theorem. Then $\forall h \in \mathbb{R}^d$, we have

$$h = \sum_{i=1}^{d} h_i \epsilon_i$$

$$\Rightarrow h^T A h = \sum_{i,j} h_i h_j \, \epsilon_i^T A \epsilon_j$$

$A\epsilon_j = \lambda_j(A)\epsilon_j$

$$= \sum_{i,j} h_i h_j \, \lambda_j(A) \, \epsilon_i^T \epsilon_j$$

$\epsilon_i^T \epsilon_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$

$$= \sum_{i=1}^{d} h_i^2 \lambda_i(A)$$

$\lambda_i(A) \geq \frac{\alpha}{2}$

$$\geq \frac{\alpha}{2} \sum_{i=1}^{d} h_i^2$$

$$= \frac{\alpha}{2} \|h\|^2.$$

Which concludes the proof.

Exercise 2 (One step of projected gradient descent)

Consider a closed convex set $\Theta \subset \mathbb{R}^d$ and a differentiable convex function $f : \Theta \to \mathbb{R}$.
For $x \in \Theta$, denote

$$x^+ := \pi_\Theta \left(x - \eta \nabla f(x) \right),$$

where π_Θ is the projection onto Θ and $\eta > 0$.

Prove that, if there exist $x^* \in \Theta$ such that $f(x^*) = \min_{x \in \Theta} f(x)$, then

$$f(x) - f(x^*) \leq \frac{\|x - x^*\|^2 - \|x^+ - x^*\|^2}{2\eta} + \frac{\eta}{2} \|\nabla f(x)\|^2$$

Solution.

By convexity of f (and the fact that, for any convex differentiable function the gradient $\nabla f(x)$ is a (the unique) subgradient of f at x) we obtain that

$$f(x^*) \geq f(x) + \nabla f(x)^T (x^* - x)$$

$$\Longleftrightarrow \quad f(x) - f(x^*) \leq \nabla f(x)^T (x - x^*). \quad (*_1)$$

Now observe that, since $x^* \in \Theta \Rightarrow x^* = \pi_\Theta(x^*)$, and since π_Θ is 1-Lipschitz:

$$\|x^+ - x^*\| = \|\pi_\Theta(x - \eta \nabla f(x)) - \pi_\Theta(x^*)\|$$

$$\leq \|x - \eta \nabla f(x) - x^*\|$$

Expanding the last term, we obtain:

$$\|x^+ - x^*\|^2 \leq \|x - \eta \nabla f(x) - x^*\|^2$$

$$= \|x - x^*\|^2 - 2\eta \nabla f(x)^T (x - x^*) + \eta^2 \|\nabla f(x)\|^2$$

Reordering terms $\Rightarrow$

$$\nabla f(x)^T (x - x^*) \underset{(*_2)}{\leq} \frac{\|x - x^*\|^2 - \|x^+ - x^*\|^2}{2\eta} + \frac{\eta}{2} \|\nabla f(x)\|^2$$

Plugging $(*_2)$ into $(*_1)$ implies the desired result.

$\square$