C6.2/B2. Continuous Optimization

Problem Sheet 2 – Solution of Problem 6

Problem 6. (to be covered in problem class/revision) Given a symmetric $n \times n$ matrix M, show that

$$\lambda_{\min}(M) \leq \frac{s^T M s}{\|s\|^2} \leq \lambda_{\max}(M), \quad \text{for all } s \in \Re^n, \tag{1}$$

where $\lambda_{\min}(M)$ and $\lambda_{\max}(M)$ are the smallest and largest eigenvalues of M, respectively.

Let $f : \Re^n \longrightarrow \Re$ be twice continuously differentiable. Then the gradient ∇f is Lipschitz continuous, namely, there exists $L \geq 0$ such that

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \quad \text{for all } x,\, y \in \Re^n, \tag{2}$$

if and only if the Hessian $\nabla^2 f$ is uniformly bounded above, that is,

$$\|\nabla^2 f(x)\| \leq L, \quad \text{for all } x \in \Re^n. \tag{3}$$

Comment: Note that this problem is in two parts. It is meant to explain some statements in the proof of Theorem 10 (global convergence of damped Newton method. For the second part, it is sufficient to prove just one of the implications.)

Solution. Equation (1) is the so-called Rayleigh-quotient inequality. As M is symmetric, it is diagonalizable and there exists an orthonormal matrix Q and a diagonal matrix of its (real) eigenvalues Λ such that $M = Q\Lambda Q^T$. Then we have that

$$\frac{s^T M s}{\|s\|^2} = \frac{(Q^T s)\Lambda Q^T s}{\|s\|^2} = \frac{(Q^T s)\Lambda Q^T s}{s^T Q Q^T s} = \frac{(Q^T s)\Lambda Q^T s}{\|Q^T s\|^2} = \frac{v^T \Lambda v}{\|v\|^2},$$

where $v = Q^T s$, and where we also used that Q is orthonormal and the definition of Euclidean norm. Since $\Lambda = (\lambda_i)_{i=\overline{1,n}}$, we deduce

$$\frac{s^T M s}{\|s\|^2} = \frac{\sum_{i=1}^n \lambda_i v_i^2}{\|v\|^2},$$

where $v = (v_1, \ldots, v_n)$. As $v_i^2 \geq 0$ for all i, clearly $(\min_i \lambda_i)\|v\|^2 \leq \sum_{i=1}^n \lambda_i v_i^2 \leq (\max_i \lambda_i)\|v\|^2$ and so (1) holds. The equality in LHS inequality in (1) is attained for s =eigenvector corresponding to left-most (min) eigenvalue; similarly, equality in the RHS inequality is attained for the eigenvector corresponding to the largest/right-most eigenvalue.

For the second part, let us first recall Taylor's theorem for vector functions that for any x and y in $\Re^n$,

$$\nabla f(y) = \nabla f(x) + \int_0^1 \nabla^2 f(x + t(y - x))(y - x)dt = \nabla f(x) + \left(\int_0^1 \nabla^2 f(x + t(y - x))dt\right)(y - x). \tag{4}$$

To show that (3) implies (2), use (4) to deduce the first equality below, namely, that for any x, y in $\Re^n$,

$$\begin{aligned} \|\nabla f(y) - \nabla f(x)\| &= \left\|\left(\int_0^1 \nabla^2 f(x + t(y - x))dt\right) \cdot (y - x)\right\| \leq \left\|\int_0^1 \nabla^2 f(x + t(y - x))dt\right\| \cdot \|y - x\| \\ &\leq \left(\int_0^1 \|\nabla^2 f(x + t(y - x))\|dt\right)\|y - x\| \leq \left(\int_0^1 L dt\right)\|y - x\| = L\|y - x\|, \end{aligned}$$

where the first and second inequality above follow from the (Euclidean) matrix-norm inequality $\|Ax\| \leq \|A\| \cdot \|x\|$ for any A and x, and from integral properties, while in the last inequality above we used (3).

For the converse implication, we need a slightly different variant of (4). Namely, any $x \in \Re^n$ and any $s \in \Re^n$, and any scalar $\alpha > 0$, Taylor's theorem for vector functions gives

$$\nabla f(x + \alpha s) = \nabla f(x) + \int_0^\alpha \nabla^2 f(x + ts) s dt = \nabla f(x) + \left(\int_0^\alpha \nabla^2 f(x + ts) dt \right) \cdot s. \tag{5}$$

Now fix x and s. Then (5) implies

$$\left\| \left(\int_0^\alpha \nabla^2 f(x + ts) dt \right) \cdot s \right\| = \|\nabla f(x + \alpha s) - \nabla f(x)\| \leq L\alpha \|s\|,$$

where in the last inequality we used (2). Dividing the above displayed inequality by α, we obtain

$$\frac{1}{\alpha} \left\| \left(\int_0^\alpha \nabla^2 f(x + ts) dt \right) \cdot s \right\| \leq L\|s\|.$$

Letting $\alpha \longrightarrow 0$, and using Leibniz integral rule, we deduce

$$\left\| \nabla^2 f(x) s \right\| \leq L\|s\|,$$

and so $\left\| \nabla^2 f(x) s \right\| / \|s\| \leq L$. By definition of the l_2-norm of a matrix, the latter implies that (3) holds. **End of solution.**