P4 - Optimization Questions

ST446 Distributed Computing for Big Data

Andreas Stöffelbauer

5 April 2021

P4-i) Convex Optimization Problem

$$f(x) = c + g^T x + \frac{1}{2} x^T H x$$

The gradient and the Hessian are

$$\nabla f(x) = g + Hx$$

$$\nabla^2 f(x) = H$$

Function $f(x)$ is convex if and only if either (all are equivalent) the Hessian is positive semi-definite, or $xHx \geq 0$ for all x, or all eigenvalues of the Hessian are non-negative. This is a sufficient condition.

A necessary but not sufficient condition is that all $a_1, ..., a_{n-2}$ and $tr(A) = (a_{n-1}, a_n)$ are non-negative. In fact, $a_1, ..., a_{n-2}$ are the eigenvalues of D (see part ii).

Note that xHx using only the first $n-2$ dimensions leads to $a_1 x_1^2 + a_2 x_2^2 + ... + a_{n-2} x_{n-2}^2$ which is non-negative if $a_1, ..., a_{n-2} \geq 0$. So what remains is the condition on B. In fact, another necessary condition for this is that all pivots of B must be non-negative whereas the sufficient condition is that the eigenvalues of B must be non-negative.

P4-ii) Gradient Descent Algorithm

A (local) minimum of $f(x)$ can be found by repeatedly applying the following gradient descent update using the first derivative $\nabla f(x) = g + Hx$ and a step size parameter η.

$$x^{t+1} = x^t - \eta \nabla f(x^t) = x^t - \eta(g + Hx^t)$$

P4- iii) β-Smoothness

Definition of β-smoothness: $\nabla^2 f(x) \preccurlyeq \beta I$.

Since $\nabla^2 f(x) = H$, function $f(x)$ is β-smooth if and only if $H \preccurlyeq \beta I$. That is, if and only if all the eigenvalues of the Hessian of $f(x)$ are at most equal to β.

The eigenvalues of the Hessian can be calculated by solving

$$0 = det(H - \lambda I_n) = det(D - \lambda I_{n-2})\, det(B - \lambda I_2)$$

This is due to the fact that H is a diagonal block matrix. We can solve this by independently finding the eigenvalues of D and B.

The eigenvalues of a diagonal matrix are its diagonal elements. Therefore, the eigenvalues of matrix D are simply $a_1, a_2, ..., a_{n-2}$.

The eigenvalues of of B can be found by solving

$$0 = det(B - \lambda I) = det \begin{pmatrix} a_{n1} - \lambda & b \\ b & a_n - \lambda \end{pmatrix} = (a_{n-1} - \lambda)(a_n - \lambda) - b^2$$

which is a quadratic equation with two solutions λ_{n-1} and λ_n (not necessarily unique). Since B is symmetric, we know that these eigenvalues indeed exist (that is, they are real).

By definition, function $f(x)$ is β-smooth with

$$\beta \leq max\{a_1, ..., a_{n-2}, \lambda_{n-1}, \lambda_n\}$$

P4- iv) Upper Bound

The obvious eigenvalues of H are $\lambda_1 = 6, \lambda_2 = 4$. In addition, we need the two eigenvalues of the matrix B at the center of H. Those eigenvalues can be calculated by solving

$$0 = det(B - \lambda I) = det \begin{pmatrix} 4 - \lambda & 2 \\ 2 & 1 - \lambda \end{pmatrix}$$

That is, $0 = det(B - \lambda I) = (4 - \lambda)(1 - \lambda) - 2^2 = \lambda^2 - 5\lambda = \lambda(\lambda - 5)$.

This leads to the eigenvalues $\lambda_3 = 0$ and $\lambda_4 = 5$.

An upper bound on the difference of $f(x^T)$ and $f(x^*)$ based on the β-smoothness condition of $f(x)$ using step size $\eta = \frac{1}{\beta} = \frac{1}{6}$ satisfies

$$f(x^T) - f(x^*) \leq 2\beta ||x^0 - x^*||^2 \frac{1}{T-1} = 12||x^0 - x^*||^2 \frac{1}{T-1}$$

which decreases with a factor $\frac{1}{T-1}$.

In addition, we can also determine α-strong convexity.

Definition:$\nabla^2 f(x) \succcurlyeq \alpha I$, or in words: $f(x)$ is α-strongly convex if and only if the smallest eigenvalue of $f(x)$ is greater or equal to α.

We can see that $f(x)$ is not strongly convex since strong convexity requires $\alpha > 0$ and $\min_i\{\lambda_i\} = 0$.

An upper bound on the difference of $f(x^T)$ and $f(x^*)$ based on both β-smoothness and α-strong convexity with step size $\eta = \frac{2}{\alpha+\beta} = \frac{2}{6}$ would be

$$f(x^T) - f(x^*) \leq \frac{\beta}{2} ||x^0 - x^*||^2 \exp(-4\frac{\alpha}{\alpha + \beta}(T-1)) = 3||x^0 - x^*||^2$$

which is independent of T and therefore suggests that the gradient descent will never converge. In fact, $f(x)$ is only convex and not strictly convex since the Hessian is only positive semi-definite and not positive definite. That means, the gradient descent can get stuck in or converge to a saddle point.

P4- v) Upper Bound #2

The eigenvalues of H are $\lambda_1 = 6, \lambda_2 = 4$ plus the two eigenvalues of the matrix B, which can be obtained as follows.

$0 = det(A - \lambda I) = (4 - \lambda)(4 - \lambda) - 2^2 = \lambda^2 - 8\lambda + 12$

Solving this equation yields $\lambda_3 = 2$ and $\lambda_4 = 6$.

We can now see that $f(x)$ is both β-smooth with $\beta = \max_i\{\lambda_i\} = 6$ and α-strongly convex with $\alpha = \min_i\{\lambda_i\} = 2$.

An upper bound on the difference of $f(x^T)$ and $f(x^*)$ using only the β-smoothness condition and a step size of $\eta = \frac{1}{\beta} = \frac{1}{6}$ is

$$f(x^T) - f(x^*) \leq 2\beta||x^0 - x^*||^2 \frac{1}{T-1} = 12||x^0 - x^*||^2 \frac{1}{T-1}$$

Using not only the β-smoothness but also the α-strongly convex condition, an upper bound using step size $\eta = \frac{2}{\alpha+\beta} = \frac{1}{4}$ is

$$f(x^T) - f(x^*) \leq \frac{\beta}{2}||x^0 - x^*||^2 \exp(-4\frac{\alpha}{\alpha+\beta}(T-1)) = 3||x^0 - x^*||^2 \exp(-T+1)$$

This upper bound converges to zero exponentially with a factor of $\exp(-T+1)$ whereas the upper bound based on smoothness decreases more slowly with a factor of $\frac{1}{T-1}$.