Assignment 2, Part 4

Anya P. Conti

April 10, 2019

1) The necessary and sufficient conditions for optimization to be convex are that the Hessian must be positive semi-definite.

$$m^T(\nabla^2 f(\mathbf{x}))m \geq 0 \quad \forall \mathbf{x} \neq 0 \qquad \text{all eigen values are non-negative} \tag{1}$$

$$\begin{aligned}
f(\mathbf{x}) &= c + \mathbf{g}^\top \mathbf{x} + \frac{1}{2}\mathbf{x}^\top \mathbf{H}\mathbf{x} && (2)\\
&= c + \sum_{i=1}^{n} g_i x_i + 0.5 \sum_{i=1}^{n}\sum_{j=1}^{n} H_{ij} x_i x_j && (3)\\
\nabla f(\mathbf{x}) &= \mathbf{g} + 0.5\mathbf{H}\mathbf{x} + 0.5\mathbf{H}^\top \mathbf{x} && (4)\\
&= \mathbf{g} + \mathbf{H}\mathbf{x} \qquad \text{Since } \mathbf{H} \text{ is symmetric} && (5)\\
\nabla^2 f(\mathbf{x}) &= \mathbf{H} && (6)
\end{aligned}$$

Thus, the eigen-values λ of $\mathbf{H}$ must be non-negative. To solve for λ:

$$\begin{aligned}
det(\mathbf{H} - \lambda I) &= 0 && (7)\\
\left(\prod_{i=1}^{n-2}(a_i - \lambda)\right)\left[(a_{n-1} - \lambda)(a_n - \lambda) - (b)^2\right] &= 0 && (8)
\end{aligned}$$

To solve the second part:

$$\begin{aligned}
0 &= \left[(a_{n-1} - \lambda)(a_n - \lambda) - (b)^2\right] && (9)\\
0 &= a_{n-1}a_n - a_{n-1}\lambda - a_n\lambda + \lambda^2 - b^2 && (10)\\
0 &= \lambda^2 + (-a_{n-1} - a_n)\lambda + (a_{n-1}a_n - b^2) && (11)\\
\lambda &= 0.5(a_{n-1} + a_n) \pm 0.5\sqrt{(-a_{n-1} - a_n)^2 - 4(a_{n-1}a_n - b^2)} && (12)\\
&= 0.5(a_{n-1} + a_n) \pm 0.5\sqrt{a_{n-1}^2 - 2a_{n-1}a_n + a_n^2 - 4a_{n-1}a_n - 4b^2} && (13)\\
&= 0.5(a_{n-1} + a_n) \pm 0.5\sqrt{a_{n-1}^2 - 6a_{n-1}a_n + a_n^2 - 4b^2} && (14)
\end{aligned}$$

Thus, we get:

$$\lambda = a_1, a_2, ..., a_{n-2}, 0.5(a_{n-1} + a_n) \pm 0.5\sqrt{a_{n-1}^2 - 6a_{n-1}a_n + a_n^2 - 4b^2} \tag{15}$$

Since as already mentioned, the eigen values have to be non-negative, then $\lambda \geq 0$ so for each value of λ, that condition has to be true: $a_1 \geq 0, a_2 \geq 0, ...$ etc.

2) The update equation for the gradient descent algorithm with a constant step size η for the optimization problem (1) would be the following:

$$w_{t+1} = w_t - \eta \nabla f(w_t) \tag{16}$$

$$w_{t+1} = w_t - \eta(g + Hw_t) \tag{17}$$

3) f is β-smooth if $||\nabla f(y) - \nabla f(x)|| \leq \beta ||y - x||$ thus:

$$\frac{||\nabla f(y) - \nabla f(x)||}{||y - x||} \leq \beta I \tag{18}$$

$$\nabla^2 f \leq \beta I \tag{19}$$

$$\nabla^2 f - \beta I \leq 0 \tag{20}$$

$$det(\nabla^2 f - \lambda I) = 0 \tag{21}$$

$$det(nabla^2 f - \beta I) \leq det(\nabla^2 f - \lambda I) \tag{22}$$

$$\lambda I \leq \beta I \tag{23}$$

So f is β-smooth if all eigen values are less than or equal to β.

4)

$$det(H - \lambda I) = 0 \tag{24}$$

$$(6 - \lambda)[(4 - \lambda)(1 - \lambda) - 4](4 - \lambda) = 0 \tag{25}$$

$$(6 - \lambda)[4 - \lambda - 4\lambda + \lambda^2 - 4](4 - \lambda) = 0 \tag{26}$$

$$(6 - \lambda)[\lambda^2 - 5\lambda](4 - \lambda) = 0 \tag{27}$$

$$(6 - \lambda)(\lambda)(\lambda - 5)(4 - \lambda) = 0 \tag{28}$$

$$\lambda = 0, 4, 5, 6 \tag{29}$$

Thus this is convex (but not strongly convex) and with a smooth gradient, which means that the best upperbound that can be guaranteed for $f(\mathbf{x}_T) - f(\mathbf{x}*)$ is as follows:

$$f(\mathbf{x}_T) - f(\mathbf{x}*) \leq \frac{2\beta||x_0 - x*||^2}{t - 1} \tag{30}$$

5)

$$det(H - \lambda I) = 0 \tag{31}$$
$$(6 - \lambda)[(4 - \lambda)(4 - \lambda) - 4](4 - \lambda) = 0 \tag{32}$$
$$(6 - \lambda)[16 - 4\lambda - 4\lambda + \lambda^2 - 4](4 - \lambda) = 0 \tag{33}$$
$$(6 - \lambda)[\lambda^2 - 8\lambda + 12](4 - \lambda) = 0 \tag{34}$$
$$(6 - \lambda)(\lambda - 6)(\lambda - 2)(4 - \lambda) = 0 \tag{35}$$
$$\lambda = 2, 4, 6 \tag{36}$$

Thus this is strongly convex with a smooth gradient, which means that the best upperbound that can be guaranteed for $f(\mathbf{x}_T) - f(\mathbf{x}*)$ is as follows:

$$f(\mathbf{x}_T) - f(\mathbf{x}*) \leq \frac{\beta}{2}||x_0 - x*||^2 e^{-4\frac{\alpha}{\alpha-\beta}t} \tag{37}$$