Introduction to Optimization Homework (4)

Dimitri Tabatadze · Monday 25-03-2024

Problem 4.1:

a) *Proof*: Let $P \subset \mathbb{N}_0$ and $Q \subset \mathbb{N}_0 \setminus P$ be the indices of the infinite subsequences of $\{x^k\}_{k \in \mathbb{N}_0}$ that converges to $\bar{x}$ and $\tilde{x}$ respectively.
W.L.O.G. assume that $f(\bar{x}) < f(\tilde{x})$. Since $\{x^k\}_{k \in \mathbb{N}_0}$ is strictly monotonously decreasing, by convergence, there will exist $i \in \mathbb{N}$ such that
$$f(x^{i-1}) > f(\tilde{x}) \geq f(x^i)$$
and again by monotonousity, $f(\tilde{x}) > f(x^j) \forall j > i$ and $|f(\tilde{x}) - f(x^j)| > |f(\tilde{x}) - f(x^{j+1})|$, i.e. after i, difference between $f(\tilde{x})$ and $f(x^j)$ grows as $j \to \infty$. This means that the subsequence converging to $\tilde{x}$ would have to be before x^j and therefore have a finite size, which is a contradiction (a finite sequence can't converge to anything). $\square$

b) Since $f(x^{k+1}) < f(x^k)$ (strictly monotonously decreasing), any subsequence would converge to a value smaller than the value of f at any x^k. **TODO:**

Graded Homework Assignment

Problem IV.1:

Proof: Let $Q = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$, $D = \begin{pmatrix} \frac{1}{a} & 0 \\ 0 & \frac{1}{c} \end{pmatrix}$, then $D^{\frac{1}{2}} = \begin{pmatrix} \frac{1}{\sqrt{a}} & 0 \\ 0 & \frac{1}{\sqrt{c}} \end{pmatrix}$ and $D^{\frac{1}{2}} Q D^{\frac{1}{2}} = \begin{pmatrix} 1 & \frac{b}{\sqrt{ac}} \\ \frac{b}{\sqrt{ac}} & 1 \end{pmatrix}$.
Now we can write
$$\det(Q - \lambda I) = (a - \lambda)(c - \lambda) - b^2 = \lambda^2 - (a + c)\lambda + ac - b^2$$
which gives eighenvalues $\frac{a+c-\sqrt{(a-c)^2+4b^2}}{2}$ and $\frac{a+c+\sqrt{(a-c)^2+4b^2}}{2}$. And
$$\det\left(D^{\frac{1}{2}} Q D^{\frac{1}{2}} - \lambda I\right) = (1 - \lambda)^2 - \left(\frac{b}{\sqrt{ac}}\right)^2 - b^2$$
with eighenvalues $1 - \frac{b}{\sqrt{ac}}$ and $1 + \frac{b}{\sqrt{ac}}$.
We need to show that
$$\frac{\cancel{2}\left(a + c + \sqrt{(a-c)^2 + 4b^2}\right)}{\cancel{2}\left(a + c - \sqrt{(a-c)^2 + 4b^2}\right)} \geq \frac{1 + \frac{b}{\sqrt{ac}}}{1 - \frac{b}{\sqrt{ac}}}$$
Let $x = a + c, y = \sqrt{(a-c)^2 + 4b^2}, z = \frac{b}{\sqrt{ac}}$. We know that $x \pm y > 0$ and it's easy to show that $1 \pm d > 0$. Now we need to show that

$$
\begin{aligned}
(x+y)(1-z) \geq (x-y)(1+z) &\iff x+y-xz-yz \geq x-y+xz-yz \\
&\implies 2y \geq 2xz \\
&\implies \sqrt{(a-c)^2+4b^2} \geq (a+c)\frac{b}{\sqrt{ac}} \\
&\implies ac\left((a-c)^2+4b^2\right) \geq (a+c)^2 b^2 \\
&\implies ac(a-c)^2 \geq (a+c)^2 b^2 - 4acb^2 \\
&\implies ac(a-c)^2 \geq (a-c)^2 b^2 \\
&\implies ac \geq b^2
\end{aligned}
$$

which is true since $\det(Q) = ac - b^2 > 0$. □

Problem IV.2:

a) - $\nabla \hat{f}(y) = \nabla(f(T(y))) = \nabla(f(By+b)) = B^T \nabla f(By+b)$
   - $H_{\hat{f}}(y) = B^T H_f(By+b)B$ from the slides.

b) Let $x_g = x + s$. Then,

$$
\begin{aligned}
s &= T(y_g) - x \\
&= T\left(y - \alpha\nabla\hat{f}(y)\right) - x \\
&= B\left(y - \alpha\nabla\hat{f}(y)\right) + b - x \\
&= By + b - \alpha B\nabla\hat{f}(y) - x \\
&= T(y) - \alpha BB^T\nabla f(x) - x \\
&= \alpha BB^T \nabla f(x) \\
&\Downarrow \\
x_g &= x + \alpha BB^T \nabla f(x)
\end{aligned}
$$

and

$$
\begin{aligned}
T^{-1}(x_g) &= B^{-1}(x_g - b) \\
&= B^{-1}(x + s - b) \\
&= B^{-1}(x + \alpha BB^T \nabla f(x) - b) \\
&= B^{-1}(x - b) + \alpha B^T \nabla f(x) \\
&= y + \alpha \nabla \hat{f}(y) \\
&= y_g.
\end{aligned}
$$

This means that a step $-\alpha\nabla\hat{f}(x)$ in the transformed space can be seen as a step in the original space with the search direction $T(y_g) - x$. Now the linear system

$$
\begin{aligned}
Ms &= -\nabla f(x) \\
&\Downarrow \\
M(-\alpha BB^T \nabla f(x)) &= -\nabla f(x) \\
&\Downarrow \\
M\alpha BB^T &= I \\
&\Downarrow \\
M &= (\alpha BB^T)^{-1} \\
&= \frac{1}{\alpha} B^{-T} B^{-1}
\end{aligned}
$$

So M is positive semi-definite

c)

$$
\begin{gathered}
T(y_g) = x_g \\
\Downarrow \\
T\left(y - \alpha \nabla \hat{f}(y)\right) = x - \alpha \nabla f(x) \\
\Downarrow \\
T\left(T^{-1}(x) - \alpha \nabla \hat{f}(T^{-1}(x))\right) = x - \alpha \nabla f(x) \\
\Downarrow \\
T(T^{-1}(x) - \alpha B^T \nabla f(T(T^{-1}(x)))) = x - \alpha \nabla f(x) \\
\Downarrow \\
T(T^{-1}(x) - \alpha B^T \nabla f(x)) = x - \alpha \nabla f(x) \\
\Downarrow \\
B(B^{-1}(x - b) - \alpha B^T \nabla f(x)) + b = x - \alpha \nabla f(x) \\
\Downarrow \\
x - \alpha B B^T \nabla f(x) = x - \alpha \nabla f(x) \\
\Downarrow \\
B B^T \nabla f(x) = \nabla f(x) \\
\Downarrow \\
B B^T = I
\end{gathered}
$$

Therefore B must be an *orthogonal* matrix

d)