Continuous Optimization: Assignment 5

Due on May 28, 2024

Honglu Ma **Hiroyasu Akada** **Mathivathana Ayyappan**

Exercise 1

Consider quadratic function $f(x) = \frac{1}{2}x^\top Qx + b^\top x$ where Q is a symmetric positive definite matrix. Using Newton's method we obtain

$$\begin{aligned} x^{(1)} &= x^{(0)} - (\nabla^2 f(x^{(0)}))^{-1}\,\nabla f(x^{(0)}) \\ &= x^{(0)} - Q^{-1}(Qx^{(0)} + b) \\ &= x^{(0)} - Q^{-1}Qx^{(0)} + Q^{-1}b \\ &= x^{(0)} - x^{(0)} + Q^{-1}b \\ &= Q^{-1}b \end{aligned}$$

$\nabla f(x^{(1)}) = -QQ^{-1}b + b = 0 \Rightarrow x^{(1)}$ is the global minimizer

Exercise 2

(a)

$Q \in \mathbb{S}_{++}(n)$ with eigenvectors $(v_1, \ldots, v_n)$. We want to show $v_i^\top Qv_j = 0$ for any $i \neq j$. We know $Q = U\Lambda U^\top$ where

$$U = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix}$$

U is an orthogonal matrix i.e. $UU^\top = I$.
For any $i \neq j$, we have

$$\begin{aligned} v_i^\top Qv_j &= v_i^\top U\Lambda U^\top v_j \\ &= v_i^\top \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix} \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix}^\top v_j \\ &= \begin{pmatrix} 0 & \cdots & v_i^\top v_i & \cdots & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ v_j^\top v_j \\ \vdots \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & \cdots & \lambda_i v_i^\top v_i & \cdots & 0 \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ v_j^\top v_j \\ \vdots \\ 0 \end{pmatrix} \\ &= 0 \end{aligned}$$

(b)

$\{d^{(0)}, \cdots, d^{(k)}\}$ are conjugate directions w.r.t. Q. We have $d_i^\top Qd_j = 0$ for any $i \neq j$.

Proof by Contradiction:
Assume that $\{d^{(0)}, \cdots, d^{(k)}\}$ is not a linearly independent set. Then there exists a non-zero vector d_j such that d_j can be expressed as a linear combination of some other vectors in the set i.e. $d_j = \sum_{i \neq j} \delta_i d_i$. Take d_m that is part of the linear combination. We have

$$\begin{aligned}
d_j^\top Q d_m &= \left(\sum_{i \neq j} \delta_i d_i \right)^\top Q d_m \\
&= \delta_m d_m^\top Q d_m + \left(\sum_{i \neq j, i \neq m} \delta_i d_i \right)^\top Q d_m \\
&= \delta_m d_m^\top Q d_m && \text{Property of Conjugate Directions} \\
&\neq 0 && Q \in \mathbb{S}_{++}(n)
\end{aligned}$$

which is a contradiction.

(c)

For all $i = 0, \cdots, k-1$, we have

$$\begin{aligned}
\langle d^{(i)}, \nabla f(x^{(k)}) \rangle &= \langle d^{(i)}, Qx^{(k)} - b \rangle \\
&= \langle Qx^{(k)}, d^{(i)} \rangle - \langle b, d^{(i)} \rangle \\
&= \langle Q \left(x^{(i+1)} + \sum_{j=i+1}^{k-1} \tau_j d^{(j)} \right), d^{(i)} \rangle - \langle b, d^{(i)} \rangle \\
&= \langle Qx^{(i+1)}, d^{(i)} \rangle + \sum_{j=i+1}^{k-1} \langle \tau_j d^{(j)}, Qd^{(i)} \rangle - \langle b, d^{(i)} \rangle \\
&= \langle Qx^{(i+1)}, d^{(i)} \rangle - \langle b, d^{(i)} \rangle && \text{Property of Conjugate Directions} \\
&= \langle Qx^{(i+1)} - b, d^{(i)} \rangle \\
&= \langle \nabla f(x^{(i+1)}), d^{(i)} \rangle \\
&= 0 && \text{Exact line search optimal condition}
\end{aligned}$$

Exercise 3

We expand the BFGS update formula:

$$\begin{aligned}
H_{k+1} &= (I - \rho_k s^{(k)} (y^{(k)})^\top) H_k (I - \rho_k y^{(k)} (s^{(k)})^\top) + \rho_k s^{(k)} (s^{(k)})^\top \\
&= (H_k - \rho_k s^{(k)} (y^{(k)})^\top H_k)(I - \rho_k y^{(k)} (s^{(k)})^\top) + \rho_k s^{(k)} (s^{(k)})^\top \\
&= H_k - \rho_k \left(H_k y^{(k)} \right) (s^{(k)})^\top - \rho_k s^{(k)} \left((y^{(k)})^\top H_k \right) + \rho_k^2 y^{(k)} (s^{(k)})^\top \left(\left(H_k y^{(k)} \right) (s^{(k)})^\top \right) + \rho_k s^{(k)} (s^{(k)})^\top
\end{aligned}$$

by setting parentheses as above, only matrix vector multiplications are evaluated.

Exercise 4

(a)

$$
\begin{aligned}
\nabla g(z) &= A^\top \nabla f(x) \\
\nabla^2 g(z) &= A^\top \nabla^2 f(x) A
\end{aligned}
$$

(b)

Apply Newton's method to $g(z)$:

$$
\begin{aligned}
z^{(k+1)} &= z^{(k)} - (\nabla^2 g(z^{(k)}))^{-1} \nabla g(z^{(k)}) \\
&= z^{(k)} - (A^\top \nabla^2 f(x) A)^{-1} A^\top \nabla f(x) \\
&= z^{(k)} - A^{-1} \nabla^2 f(x)^{-1} A^{-\top} A^\top \nabla f(x) \\
&= z^{(k)} - A^{-1} \nabla^2 f(x)^{-1} \nabla f(x)
\end{aligned}
$$

We have such identity

$$
\begin{aligned}
x^{(k+1)} &= A z^{(k+1)} + b \\
&= A \left(z^{(k)} - A^{-1} \nabla^2 f(x)^{-1} \nabla f(x) \right) + b \\
&= A z^{(k)} - A A^{-1} \nabla^2 f(x)^{-1} \nabla f(x) + b \\
&= A z^{(k)} + b - \nabla^2 f(x)^{-1} \nabla f(x) \\
&= x^{(k)} - \nabla^2 f(x)^{-1} \nabla f(x)
\end{aligned}
$$

which is the same as applying Newton's method to $f(x)$.

(c)

Apply BFGS to $g(z)$: In case of $k = 0$, we have

$$
\begin{aligned}
z^{(1)} &= z^{(0)} - \tilde{H}_0 \nabla g(z^{(0)}) \\
&= z^{(0)} - \tilde{H}_0 A^\top \nabla f(x) \\
&= z^{(0)} - A^{-1} H_0 \nabla f(x) && H_0 = A \tilde{H}_0 A^\top
\end{aligned}
$$

now we have

$$
\begin{aligned}
x^{(1)} &= A z^{(1)} + b \\
&= A \left(z^{(0)} - A^{-1} H_0 \nabla f(x) \right) + b \\
&= A z^{(0)} - A A^{-1} H_0 \nabla f(x) + b \\
&= A z^{(0)} + b - H_0 \nabla f(x) \\
&= x^{(0)} - H_0 \nabla f(x)
\end{aligned}
$$

Exercise 4 continued on next page...

Induction hypothesis: assume $x^{(k)} = x^{(k-1)} - H_{(k-1)}\nabla f(x)$ We want to show it is also the case for $x^{(k+1)}$.
Inductive step: we have such identity

$$s^{(k)} = x^{(k+1)} - x^{(k)} = A(z^{(k+1)} - z^{(k)}) = A\tilde{s}^{(k)} \qquad \text{where } z^{(k+1)} - z^{(k)}$$
$$y^{(k)} = \nabla g(z^{(k+1)}) - \nabla g(z^{(k)}) = A^\top \nabla f(x^{(k+1)}) - A^\top \nabla f(x^{(k)}) = A^\top \tilde{y}^{(k)}$$
$$\text{where } \nabla f(x^{(k+1)}) - \nabla f(x^{(k)})$$

We can try to show by induction that this identity holds: $H_k = A\tilde{H}_k A^\top$. and the update formular for $z^{(k+1)}$ as such

$$\begin{aligned} z^{(k+1)} &= z^{(k)} - \tilde{H}_k \nabla g(z^{(k)}) \\ &= z^{(k)} - \tilde{H}_k A^\top \nabla f(x) \\ &= z^{(k)} - A^{-1} H_k \nabla f(x) \end{aligned}$$

Then

$$\begin{aligned} x^{(k+1)} &= Az^{(k+1)} + b \\ &= A\left(z^{(k)} - A^{-1} H_k \nabla f(x)\right) + b \\ &= Az^{(k)} - AA^{-1} H_k \nabla f(x) + b \\ &= Az^{(k)} + b - H_k \nabla f(x) \\ &= x^{(k)} - H_k \nabla f(x) \end{aligned}$$