Continuous Optimization: Assignment 2

Due on May 7, 2024

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

Exercise 1

(i)

The gradient of $\varphi(x)$ is given by

$$\begin{aligned}\nabla\varphi(x) &= \frac{1}{2}(A + A^\top)x - b \\ &= \frac{1}{2}(2A)x - b && A \text{ is symetric} \\ &= Ax - b\end{aligned}$$

Similarly, the Hessian of $\varphi(x)$ is given by

$$\begin{aligned}\nabla^2\varphi(x) &= (D(\nabla\varphi(x)))^\top \\ &= (A^\top)^\top = A\end{aligned}$$

Suppose $\nabla\varphi(x^\star) = 0$ and $\nabla^2\varphi(x^\star)$ is positive definite, then Theorem 6.9 shows that $x^\star$ is a local minimum of $\varphi(x)$. To meet the condition, we only need to show that $\nabla\varphi(x^\star) = 0$ because A is positive definite which meets the second part of the condition. Thus we can find the minimizer by solving $Ax = b$.

(ii)

The steepest descent direction is given by

$$\begin{aligned}d^{(k)} &= -\nabla\varphi(x^{(k)}) \\ &= -Ax^{(k)} + b = r^{(k)}\end{aligned}$$

(iii)

Let $g_k(\tau) = x^{(k)} + \tau r^{(k)}$ be the function that gives $x^{(k+1)}$ given τ at time step k. We can rewrite the objective function as $\min\limits_{\tau>0} \varphi(g_k(\tau))$ and to find the minimizer, we can solve $\frac{\partial\varphi}{\partial\tau} = 0$.

$$\begin{aligned}\frac{\partial\varphi}{\partial\tau} &= \frac{\partial g_k}{\partial\tau}\cdot\frac{\partial\varphi}{\partial g_k} \\ &= (b - Ax^{(k)})^\top(Ag_k - b) \\ &= -b^\top b + (Ax^{(k)})^\top b + b^\top Ag_k - (Ax^{(k)})^\top Ag_k \\ &= -b^\top b + (Ax^{(k)})^\top b + b^\top A(x^{(k)} + \tau(b - Ax^{(k)})) - (Ax^{(k)})^\top A(x^{(k)} + \tau(b - Ax^{(k)})) \\ &= -b^\top b + 2(Ax^{(k)})^\top b - (Ax^{(k)})^\top Ax^{(k)} + \tau(b - Ax^{(k)})^\top A(b - Ax^{(k)}) \\ &= -(b - Ax^{(k)})^\top(b - Ax^{(k)}) + \tau(b - Ax^{(k)})^\top A(b - Ax^{(k)}) = 0 \\ \Rightarrow \tau &= \frac{(b - Ax^{(k)})^\top(b - Ax^{(k)})}{(b - Ax^{(k)})^\top A(b - Ax^{(k)})}\end{aligned}$$

Exercise 2

(i)

The derivative of the function $f(x)$ at $x^{(k)}$ dotted with the direction $d^{(k)}$ can be expressed as

$$\langle \nabla f(x^{(k)})\ ,\ d^{(k)} \rangle = 2x^{(k)} * (-1) = -2x^{(k)} < 0 \text{ for every } x^{(k)} > 0.$$

This shows that $d^{(k)}$ is a descent direction.

(ii)

The descent method updates the current point $x^{(k)}$ based on the following formula:

$$x^{(k+1)} = x^{(k)} + \tau_k d^{(k)}$$

Substituting the given values, we get

$$x^{(k+1)} = x^{(k)} + 2^{-k-1} * (-1) = x^{(k)} - 2^{-k-1}$$

We need to show by induction that $x^{(k)} = 1 + 2^{-k}$ for all k.

- Base case ($k = 0$): $x^{(0)} = 1 + 2^0 = 2$, which is true.
- Inductive step: Assume $x^{(k)} = 1 + 2^{-k}$ is true for some k. We need to show that $x^{(k+1)} = 1 + 2^{-(k+1)}$.

$$\begin{aligned} x^{(k+1)} &= x^{(k)} - 2^{-k-1} \\ &= 1 + 2^{-k} - 2^{-k-1} \\ &= 1 + 2 * 2^{-k-1} - 2^{-k-1} \\ &= 1 + (2-1) * 2^{-k-1} \\ &= 1 + 2^{-(k+1)} \end{aligned}$$

By induction, $x^{(k)} = 1 + 2^{-k}$ for all k.

(iii)

As k approaches infinity, $2^{-k} \to 0$ and $1 + 2^{-k} \to 1$, showing that the sequence $x^{(k)}$ converges to 1, not 0.

(iv)

$f(x)$ has its minimum at $x = 0$. That is, while the sequence $x^{(k)}$ converges to 1, this is not the minimizer of $f(x)$. This suggests that the Wolfe's conditions might not be satisfied. We next check if the conditions holds, which can be shown as

$$\begin{aligned} &\langle \nabla f(\overline{x^{(k)}} + \tau_k d^{(k)})\ ,\ d^{(k)} \rangle \leq \eta \langle \nabla f(\overline{x^{(k)}})\ ,\ d^{(k)} \rangle \text{ for some } \eta \in (\gamma, 1) \\ &\Rightarrow \langle 2(\overline{x^{(k)}} + \tau_k d^{(k)})\ ,\ d^{(k)} \rangle \leq \eta \langle 2(\overline{x^{(k)}})\ ,\ d^{(k)} \rangle \\ &\Rightarrow \langle 2(\overline{x^{(k)}} + \tau_k(-1))\ ,\ (-1) \rangle \leq \eta \langle 2(\overline{x^{(k)}})\ ,\ (-1) \rangle \\ &\Rightarrow -2(\overline{x^{(k)}} + \tau_k(-1)) \leq -2\eta(\overline{x^{(k)}}) \\ &\Rightarrow \overline{x^{(k)}} - \tau_k \geq \eta \overline{x^{(k)}} \\ &\Rightarrow (1-\eta)\overline{x^{(k)}} \geq \tau_k \end{aligned}$$

Exercise 2 continued on next page...

This inequality does not hold. This is because as k approaches infinity, $\tau_k \to 0$ but $(1-\eta)\overline{x^{(k)}} \to (1-\eta)$ that is larger than 0 since $\eta < 0$. Thus, the Wolfe's conditions are not be satisfied.

Exercise 3

(a)

To show that f has a global minimizer at (a, a^2) we need to show that $f(a, a^2) \le f(x_1, x_2)$ for all (x_1, x_2).

$$f(a, a^2) - f(x_1, x_2) = 0 - (a - x_1)^2 - b(x_2 - x_1^2)^2 \le 0$$

(b)

We have the following gradient of f:

$$\nabla f(x_1, x_2) = \begin{pmatrix} 2(x_1 - a) + 4b(x_1^2 - x_2)x_1 \\ 2b(x_2 - x_1^2) \end{pmatrix}$$

observe that the gradient can be arbitarily large towards infinity, thus there does not exist an upper bound for the gradient.

From the Lipschitz condition $||\nabla f(x) - \nabla f(y)|| \le L||x - y|| \Rightarrow \frac{||\nabla f(x) - \nabla f(y)||}{||x-y||} \le L$ we should show $\exists x, y$ such that $\frac{||\nabla f(x) - \nabla f(y)||}{||x-y||}$ is unbounded, intead of showing that the gradient is unbounded.

Correct Answer

Let $x = \begin{pmatrix} t \\ 0 \end{pmatrix}$ and $y = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. We have

$$||x - y|| = t$$

$$||\nabla f(x) - \nabla f(y)|| = \left|\left| \begin{pmatrix} 2t + 4bt^3 \\ -2bt^2 \end{pmatrix} \right|\right| = 2bt^2$$

thus

$$\lim_{t \to \infty} \frac{||\nabla f(x) - \nabla f(y)||}{||x - y||} = \lim_{t \to \infty} 2bt \to \infty$$

(c)

Take $x = (-1, 1), y = (0, 0), \lambda = 0.5$ we have

$$f(\lambda(x) + (1 - \lambda)y) = f(-0.5, 0.5) = 8.5 > \lambda f(x) + (1 - \lambda)f(y) = 0.5f(-1, 1) + 0.5f(0, 0) = 2.5$$

which shows that f is not convex.