1 Question 1

(4 points) Assume $f(x)$ is a continuously differentiable function, d^k is a descent direction and $\{f(x^k + \alpha d^k), \alpha > 0\}$ is bounded below. Give an example which shows that if instead we have some $0 < c_2 < c_1 < 1$ in the Wolfe condition, there may not exist any step size α which satisfies this modified Wolfe condition. (You can choose some particular c_1, c_2 in your counter example)

Solution:

Let $f(x) = x^2$, and $f'(x) = 2x$. We can easily see that it is continuously differentiable on R. Then, we let $x^k = 1$, $c_1 = 0.9$, $c_2 = 0.1$, $d^k = -1$. Note that since

$$\nabla f(x^k)^T d^k = f'(1) * (-1) = -2 < 0$$

$d^k = -1$ is legal by the definition of descent directions. Also, $f(x^k + \alpha d^k) = f(1-\alpha) = (1-\alpha)^2$ is bounded below by 0 for $\alpha > 0$ by the property of quadratic functions.

We see if there is any α value that satisfies the modified Wolfe condition.

From the first condition,

$$f(x^k + \alpha d^k) \le f(x^k) + c_1 \alpha \nabla f(x^k)^T d^k$$
$$f(1-\alpha) \le f(1) - 0.9\alpha f'(1)$$
$$(1-\alpha)^2 \le 1 - 0.9\alpha * 2 = 1 - 1.8\alpha$$
$$1 - 2\alpha + \alpha^2 \le 1 - 1.8\alpha$$
$$\alpha^2 - 0.2\alpha \le 0$$
$$\alpha(\alpha - 0.2) \le 0$$
$$0 \le \alpha \le 0.2$$

From the second condition,

$$\nabla f(x^k + \alpha d^k)^T d^k \ge c_2 \nabla f(x^k)^T d^k$$
$$-f'(1-\alpha) \ge -0.1 f'(1)$$
$$f'(1-\alpha) \le 0.1 f'(1)$$
$$2 - 2\alpha \le 0.1 * 2 = 0.2$$
$$\alpha \ge 0.9$$

However, there is no α that satisfies both $0 \le \alpha \le 0.2$ and $\alpha \ge 0.9$.

2 Question 2

(6 points) Let $f(x) = \frac{1}{2}x^T Ax + b^T x$, d^k is a descent direction, x^k is the current iteration point. Compute the exact line search step size

$$\alpha_k = \arg\min_{\alpha>0} f(x^k + \alpha d^k)$$

Then using your formula to find the exact step size for steepest descent direction.(You can assume A to be positive definite)

Solution:

According to Lemma 6.4,

$$\frac{\partial x^T Ax}{\partial x} = 2Ax$$

Then, we also have

$$\frac{\partial x^T b}{\partial x} = b$$

Hence,

$$\nabla f(x) = \frac{1}{2}2Ax + b = Ax + b$$

We let

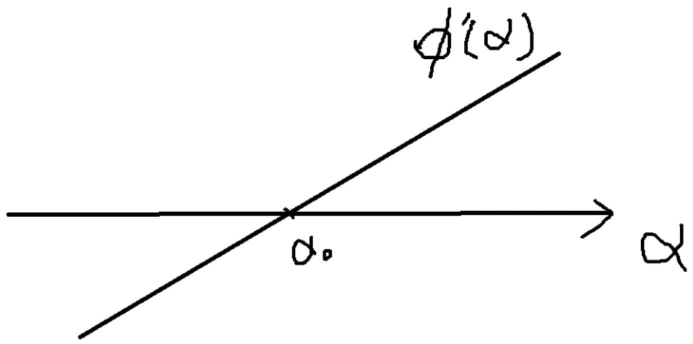
$$\Phi(\alpha) = f(x^k + \alpha d^k)$$

We have

$$\Phi'(\alpha) = \nabla f(x^k + \alpha d^k)^T d^k = [A(x^k + \alpha d^k) + b]^T d^k = (Ax^k + \alpha Ad^k + b)^T d^k$$

$$= {d^k}^T A^T d^k \alpha + (Ax^k + b)^T d^k = {d^k}^T Ad^k \alpha + \Phi'(0)$$

Note that since A is PD and $d^k \neq \mathbf{0}$, so ${d^k}^T Ad^k > 0$, and $\Phi'(0)$ does not depend on α so it's a constant term. That said, $\Phi'(\alpha)$ is a linear function with positive slope, like:



We try to get the stationary point of α: α_0:

$$\Phi'(\alpha_0) = {d^k}^T Ad^k \alpha_0 + (Ax^k + b)^T d^k = 0$$

$$d^{k^T}Ad^k\alpha_0 = -(Ax^k + b)^T d^k$$

$$\alpha_0 = -\frac{(Ax^k + b)^T d^k}{d^{k^T}Ad^k}$$

Since $\alpha_0 = -\frac{(Ax^k+b)^T d^k}{d^{k^T}Ad^k} = -\frac{\nabla f(x^k)d^k}{d^{k^T}Ad^k}$, and $\nabla f(x^k)d^k < 0$ since d^k is a descent direction, and $d^{k^T}Ad^k > 0$ since A is PD and $d^k \neq \mathbf{0}$, we know $\alpha_0 = -\frac{\nabla f(x^k)d^k}{d^{k^T}Ad^k} > 0$.

Then, we get the hessian matrix (second derivative) of $\Phi(\alpha_0)$:

$$\Phi''(\alpha) = d^{k^T}Ad^k$$

$$\Phi''(\alpha_0) = d^{k^T}Ad^k$$

For any $x \in R\ that\ x \neq 0$, $x\Phi''(\alpha_0)x = d^{k^T}Ad^k x^2$. Since A is PD, $d^{k^T}Ad^k > 0$, and $x^2 > 0$, $x\Phi''(\alpha_0)x = d^{k^T}Ad^k x^2 > 0$, we know $\Phi''(\alpha_0)$ is PD. Also, since α_0 is a stationary point, a_0 is a strict local minimizer for $\Phi(\alpha) = f(x^k + \alpha d^k)$ by Theorem 5.4. Also, since $\Phi''(\alpha) = d^{k^T}Ad^k$ is PD for any α, $\Phi(\alpha)$ is strict convex by Proposition 4.9. And we can see the domain of $\Phi(\alpha)$ is convex, since for any $\alpha_1, \alpha_2 > 0, \alpha_1 + \alpha_2 > 0$ as well. Therefore, a_0 is a global minimizer for $\Phi(\alpha) = f(x^k + \alpha d^k)$ by Theorem 3.19. Hence, $a_0 = a_k = \underset{\alpha>0}{\operatorname{argmax}} f(x^k + \alpha d^k) = -\frac{(Ax^k+b)^T d^k}{d^{k^T}Ad^k}$ since we showed that $\alpha_0 > 0$.

Then, we calculate the exact step size for the steepest descent direction:

$$d^k = -\nabla f(x^k) = -Ax^k - b$$

$$a_k = \frac{(Ax^k + b)^T d^k}{d^{k^T}Ad^k} = \frac{(Ax^k + b)^T(Ax^k + b)}{(Ax^k + b)^T A(Ax^k + b)} = \frac{\|Ax^k + b\|_2}{(Ax^k + b)^T A(Ax^k + b)}$$

3 Question 3

(4 points)

1. Let $f(x)$ be a differentiable function and Let $\{d^k\}$ be a sequence of descent directions for iteration points x^k. If x^* is any accumulation point of the sequence $\{x^k\}$, prove that $f(x^*)$ is a lower bound for the sequence $\{f(x^k)\}$.

2. Prove Corollary 8.16 in the notes. (You don't have to prove it from the beginning, you can use any results from the notes)

1. Proof:
Suppose there exists an accumulation point x^* of $\{x^k\}$, such that $f(x^*)$ is not a lower bound of $\{f(x^k)\}$ for contradiction. Then, by the definition of accumulation points, we know there is some subsequence $J \subseteq N$ such that $\lim_{j\to\infty} x^j = x^*, j \in J$. Note that since d^k is always a descent direction for x^k, we can always find a small enough α_k such that $f(x^k)$ is a decreasing sequence by Lemma 8.3. Thus, we simply assume $f(x^k)$ is a decreasing sequence, which means any subsequence of $f(x^k)$ is always decreasing as well by the definition of subsequences. That said, $\forall j \in J, \epsilon > 0\ s.t.\ j + \epsilon \in J, f(j+\epsilon) \le f(j)$. Hence, for any $j \in J$,
$$f(x^j) \ge f\left(\lim_{j\to\infty} x^j\right) = f(x^*)$$
Note that by assumption, $f(x^*)$ is not a lower bound of $\{f(x^k)\}$, so there exists $n \in N\ s.t.\ f(x^n) < f(x^*) \le f(x^j), \forall j \in J$. Since $\{x^k\}$ is decreasing, $n > j$ must hold. Also, since J is a subsequence of N, and J has infinitely many elements, we can always find $t \in J\ s.t.\ t > n$, which gives $f(x^t) \le f(x^n) < f(x^*)$ since f is decreasing, which contradicts the fact that for any $j \in J$, $f(x^j) \ge f\left(\lim_{j\to\infty} x^j\right) = f(x^*)$. Therefore, $f(x^*)$ is a lower bound of the sequence $\{f(x^k)\}$.
2. Proof:
Since d^k is a descent direction, we know $\{f(x^k)\}$ is decreasing. Then, if x^* is an accumulation point of the sequence $\{x^k\}$, we know $f(x^*)$ is a lower bound of $\{f(x^k)\}$. Then, since $\{f(x^k)\}$ is bounded below, $\|d^k\|_2 = 1$, f is differentiable and the gradient $\nabla f(x)$ is Lipschitz continuous with constant L, we know $\lim_{k\to\infty} \nabla f(x^k)^T d^k = 0$ by Theorem 8.13. That is, since $\nabla f(x^k)^T d^k < 0$, we have
$$\lim_{k\to\infty} \nabla f(x^k)^T d^k = -\lim_{k\to\infty} |\nabla f(x^k)^T d^k| = -\lim_{k\to\infty} \|\nabla f(x^k)\|_2 \|d^k\|_2 \cos(\theta_k) = 0$$
$$= -\lim_{k\to\infty} \|\nabla f(x^k)\|_2 \|d^k\|_2 \lim_{k\to\infty} \cos(\theta_k)$$

Since $\lim_{k\to\infty} \theta_k \leq \frac{\pi}{2} - \epsilon, \epsilon > 0$, and $\cos(\theta)$ is decreasing on $[0, \pi]$, we have

$$\lim_{k\to\infty} \cos(\theta_k) \geq \cos\left(\frac{\pi}{2} - \epsilon\right)$$

Hence, we have

$$\begin{aligned}
\lim_{k\to\infty} \nabla f(x^k)^T d^k = 0 = - \lim_{k\to\infty} \|\nabla f(x^k)\|_2 \|d^k\|_2 \lim_{k\to\infty} \cos(\theta_k) \\
\leq -\cos\left(\frac{\pi}{2} - \epsilon\right) \lim_{k\to\infty} \|\nabla f(x^k)\|_2 \|d^k\|_2 \\
= -\cos\left(\frac{\pi}{2} - \epsilon\right) \lim_{k\to\infty} \|\nabla f(x^k)\|_2 \\
\cos\left(\frac{\pi}{2} - \epsilon\right) \lim_{k\to\infty} \|\nabla f(x^k)\|_2 \leq 0
\end{aligned}$$

Since $\cos\left(\frac{\pi}{2} - \epsilon\right) > 0$,

$$\lim_{k\to\infty} \|\nabla f(x^k)\|_2 \leq 0$$

By the property of norms, $\|\nabla f(x^k)\|_2 \geq 0$, so $\|\nabla f(x^*)\|_2 = \lim_{k\to\infty} \|\nabla f(x^k)\|_2 =$ 0.

Therefore, $\nabla f(x^*) = 0$ by property of norms.

4 Question 4

(6 points) Let function $f : \mathbb{R}^n \to \mathbb{R}$ be a C^2-class function, which also has a Lipschitz gradient, that is,

$$||\nabla f(x) - \nabla f(y)||_2 \leq L||x - y||_2$$

for any $x, y \in \mathbb{R}$ where L is a positive number. Let d^k be a descent direction at x^k. Show that the Armijo backtracking stopping condition holds for

$$0 < \alpha \leq \frac{2(c_1 - 1)\nabla f(x^k)^T d^k}{L||d^k||_2^2}$$

Use this to give an upper bound on the number of backtracking iterations.

(Hint: you may use any results from the previous assignment)

Proof:

By Taylor Theorem,

$$f(x^k + \alpha d^k) = f(x^k) + \alpha \nabla f(x^k)^T d^k + \frac{1}{2}\alpha^2 {d^k}^T \nabla^2 f(x^k + \theta\alpha d^k) d^k, \exists 0 < \theta < 1$$

$$f(x^k + \alpha d^k) - f(x^k) = \alpha \nabla f(x^k)^T d^k + \frac{1}{2}\alpha^2 {d^k}^T \nabla^2 f(x^k + \theta\alpha d^k) d^k$$

If $\boldsymbol{{d^k}^T \nabla^2 f(x^k + \theta\alpha d^k) d^k} \geq \mathbf{0}$, then

$$f(x^k + \alpha d^k) - f(x^k) = \alpha \nabla f(x^k)^T d^k + \frac{1}{2}\alpha^2 \left|{d^k}^T \nabla^2 f(x^k + \theta\alpha d^k) d^k\right|$$

By Schwartz Inequality,

$$f(x^k + \alpha d^k) - f(x^k) \leq \alpha \nabla f(x^k)^T d^k + \frac{1}{2}\alpha^2 \left\|{d^k}^T\right\|_2 \|\nabla^2 f(x^k + \theta\alpha d^k) d^k\|_2$$

$$= \alpha \nabla f(x^k)^T d^k + \frac{1}{2}\alpha^2 \left\|{d^k}^T\right\|_2 \frac{\|\nabla^2 f(x^k + \theta\alpha d^k) d^k\|_2}{\left\|{d^k}^T\right\|_2} \left\|{d^k}^T\right\|_2$$

Since $\|\nabla^2 f(x^k + \theta\alpha d^k)\|_2 = \max_{y \neq \mathbf{0}} \frac{\|\nabla^2 f(x^k + \theta\alpha d^k) y\|_2}{\|y\|_2} \geq \frac{\|\nabla^2 f(x^k + \theta\alpha d^k) d^k\|_2}{\left\|{d^k}^T\right\|_2}$, we have

$$f(x^k + \alpha d^k) - f(x^k)$$

$$\leq \alpha \nabla f(x^k)^T d^k + \frac{1}{2}\alpha^2 \left\|{d^k}^T\right\|_2 \frac{\|\nabla^2 f(x^k + \theta\alpha d^k) d^k\|_2}{\left\|{d^k}^T\right\|_2} \left\|{d^k}^T\right\|_2$$

$$\leq \alpha \nabla f(x^k)^T d^k + \frac{1}{2}\alpha^2 \left\|{d^k}^T\right\|_2 \|\nabla^2 f(x^k + \theta\alpha d^k)\|_2 \left\|{d^k}^T\right\|_2$$

By A2q5, we have shown that $\|\nabla^2 f(y)\|_2 \leq L$ if f has Lipschitz gradient with constant L, we have

$$f(x^k+\alpha d^k)-f(x^k)\le \alpha\nabla f(x^k)^T d^k+\frac{1}{2}\alpha^2\left\|d^{k^T}\right\|_2\|\nabla^2 f(x^k+\theta\alpha d^k)\|_2\left\|d^{k^T}\right\|_2$$
$$\le \alpha\nabla f(x^k)^T d^k+\frac{1}{2}\alpha^2\left\|d^{k^T}\right\|_2^2 L$$

Therefore,

$$\begin{aligned}
&f(x^k+\alpha d^k)-f(x^k)-c_1\alpha\nabla f(x^k)^T d^k\\
&\le \alpha\nabla f(x^k)^T d^k+\frac{1}{2}\alpha^2\left\|d^{k^T}\right\|_2^2 L-c_1\alpha\nabla f(x^k)^T d^k\\
&=(1-c_1)\alpha\nabla f(x^k)^T d^k+\frac{1}{2}\alpha^2\left\|d^{k^T}\right\|_2^2 L\\
&=\alpha\left[(1-c_1)\nabla f(x^k)^T d^k+\frac{1}{2}\alpha\left\|d^{k^T}\right\|_2^2 L\right]
\end{aligned}$$

Note that since $0<\alpha\le\frac{c(c_1-1)\nabla f\left(x^k\right)^T d^k}{L\|d^k\|_2^2}$, we have

$$\begin{aligned}
&(1-c_1)\nabla f(x^k)^T d^k+\frac{1}{2}\alpha\left\|d^{k^T}\right\|_2^2 L\\
&\le (1-c_1)\nabla f(x^k)^T d^k+\frac{1}{2}\frac{2(c_1-1)\nabla f(x^k)^T d^k}{L\|d^k\|_2^2}\left\|d^{k^T}\right\|_2^2 L\\
&=(1-c_1)\nabla f(x^k)^T d^k+(c_1-1)\nabla f(x^k)^T d^k\\
&=(1-c_1)\nabla f(x^k)^T d^k-(1-c_1)\nabla f(x^k)^T d^k=0
\end{aligned}$$

Since $\alpha>0$, $(1-c_1)\nabla f(x^k)^T d^k+\frac{1}{2}\alpha\left\|d^{k^T}\right\|_2^2 L\le 0$, we have

$$f(x^k+\alpha d^k)-f(x^k)-c_1\alpha\nabla f(x^k)^T d^k\le\alpha\left[(1-c_1)\nabla f(x^k)^T d^k+\frac{1}{2}\alpha\left\|d^{k^T}\right\|_2^2 L\right]$$
$$\le 0$$
$$f(x^k+\alpha d^k)-f(x^k)\le c_1\alpha\nabla f(x^k)^T d^k$$

That said, the Armijo backtracking stopping condition is sufficient.

If $d^{k^T}\nabla^2 f\left(x^k+\theta\alpha d^k\right)d^k<0$, then

$$f(x^k+\alpha d^k)-f(x^k)=\alpha\nabla f(x^k)^T d^k-\frac{1}{2}\alpha^2\left|d^{k^T}\nabla^2 f(x^k+\theta\alpha d^k)d^k\right|$$

$$\begin{aligned}
&f(x^k+\alpha d^k)-f(x^k)-c_1\alpha\nabla f(x^k)^T d^k\\
&=\alpha\nabla f(x^k)^T d^k-\frac{1}{2}\alpha^2\left|d^{k^T}\nabla^2 f(x^k+\theta\alpha d^k)d^k\right|-c_1\alpha\nabla f(x^k)^T d^k\\
&=(1-c_1)\alpha\nabla f(x^k)^T d^k-\frac{1}{2}\alpha^2\left|d^{k^T}\nabla^2 f(x^k+\theta\alpha d^k)d^k\right|
\end{aligned}$$

Since $0<c_1<1, 0<1-c_1<0$, $\nabla f(x^k)^T d^k<0$, $\alpha>0$, and $-\frac{1}{2}\alpha^2\left|d^{k^T}\nabla^2 f(x^k+\theta\alpha d^k)d^k\right|<0$, we know

$$\begin{aligned}
&f(x^k+\alpha d^k)-f(x^k)-c_1\alpha\nabla f(x^k)^T d^k\\
&=(1-c_1)\alpha\nabla f(x^k)^T d^k-\frac{1}{2}\alpha^2\left|d^{k^T}\nabla^2 f(x^k+\theta\alpha d^k)d^k\right|<0
\end{aligned}$$

$$f(x^k + \alpha d^k) - f(x^k) < c_1 \alpha \nabla f(x^k)^T d^k$$

That said, the Armijo backtracking stopping condition is sufficient.

Let $\gamma \in (0, 1)$ be the parameter in the Armijo backtracking step. Then, we let $\alpha_0 > 0$ be the initial guess of α, and let n be the number of steps we need in the Armijo backtracking step. Then, we have

$$\alpha_0 \gamma^n \leq \frac{2(c_1 - 1)\nabla f(x^k)^T d^k}{L\|d^k\|_2}$$

Since $\alpha_0 > 0$,

$$\gamma^n \leq \frac{2(c_1 - 1)\nabla f(x^k)^T d^k}{L\|d^k\|_2 \alpha_0}$$

$$n \leq \log_\gamma \frac{2(c_1 - 1)\nabla f(x^k)^T d^k}{L\|d^k\|_2 \alpha_0}$$

Therefore, the upper bound of the number of steps is $\log_\gamma \frac{2(c_1 - 1)\nabla f(x^k)^T d^k}{L\|d^k\|_2 \alpha_0}$.