EECS 127/227AT **Fall 2020**

Solutions for Discussion 13

1 Backtracking line search

Let $f : \mathbb{R}^2 \to \mathbb{R}$ with $\mathrm{dom}(f) = \mathbb{R}^2$ be given by

$$f(x) = x_1^2 + 3x_2^2.$$

We consider backtracking line search to find the minimum of f, using the parameters $\alpha = 0.25$ and $\beta = 0.5$. Suppose the algorithm is at $x^{(k)} := \begin{bmatrix} 2 & 2 \end{bmatrix}^T$. Determine $x^{(k+1)}$.

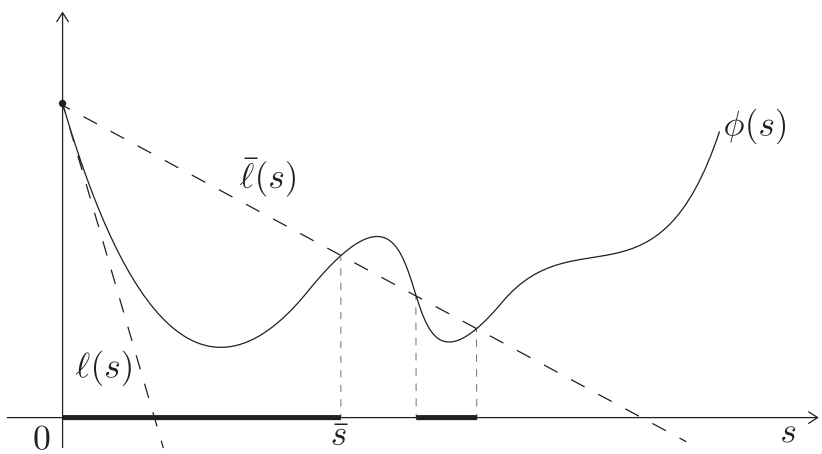

Figure 1: Backtracking line search. Here $l(s) := f(x^{(k)}) - s\|\nabla f(x^{(k)})\|_2^2$ and $\bar{l}(s) := f(x^{(k)}) - \alpha s\|\nabla f(x^{(k)})\|_2^2$. The abscissa is parametrized by s, so the graph is of $\phi(s) := f(x^{(k)} - s\nabla f(x^{(k)}))$. This is only a generic illustration—in this question the function $\phi(s)$ is convex.

Solution:

We have $\nabla f(x) = \begin{bmatrix} 2x_1 & 6x_2 \end{bmatrix}^T$. Hence backtracking search will end up comparing the function

$$\begin{aligned}
\phi(s) &:= f(x^{(k)} - s\nabla f(x^{(k)}),\ 0 \le s \le 1 \\
&= f(\begin{bmatrix} 2 \\ 2 \end{bmatrix} - s\begin{bmatrix} 4 \\ 12 \end{bmatrix}),\ 0 \le s \le 1 \\
&= f(\begin{bmatrix} 2 - 4s \\ 2 - 12s \end{bmatrix}),\ 0 \le s \le 1 \\
&= (2-4s)^2 + 3(2-12s)^2,\ 0 \le s \le 1,
\end{aligned}$$

to the function

$$\begin{aligned}
\bar{l}(s) &:= f(x^{(k)}) - \alpha s\|\nabla f(x^{(k)})\|_2^2,\ 0 \le s \le 1 \\
&= f(\begin{bmatrix} 2 \\ 2 \end{bmatrix}) - \frac{s}{4}\|\begin{bmatrix} 4 \\ 12 \end{bmatrix}\|_2^2,\ 0 \le s \le 1 \\
&= 16 - 40s,\ 0 \le s \le 1.
\end{aligned}$$

See Figure 1.
Backtracking line search starts with $s = 1$. We have

$$\phi(1) = 304 > -24 = \bar{l}(1).$$

Hence we replace s with $s\beta = 0.5$. We then have

$$\phi(0.5) = 48 > -4 = \bar{l}(0.5).$$

Hence we replace s with $s\beta = 0.25$. Now we have

$$\phi(0.25) = 4 < 6 = \bar{l}(0.25),$$

i.e. the Armijo condition is now satisfied. Therefore backtracking line search will set

$$x^{(k+1)} = x^{(k)} - 0.25\nabla f(x^{(k)}) = \begin{bmatrix} 2 \\ 2 \end{bmatrix} - 0.25 \begin{bmatrix} 4 \\ 12 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

2 Pure Newton method

Let $a : \mathbb{R}_+ \to \mathbb{R}$ satisfy the following properties:

(I) a is a continuous nondecreasing function on $\mathbb{R}_+$;

(II) $a(0) = 0$;

(III) a is uniformly bounded above on $\mathbb{R}_+$; i.e., there exists some constant $K < \infty$ such that $a(x) \leq K$ for all $x \in \mathbb{R}_+$;

(IV) a is differentiable on $\mathbb{R}_{++}$ with $\lim_{x \downarrow 0} a'(x) = 0$;

and

(V)

$$\lim_{x \to \infty} \frac{a(x)}{x a'(x)} = \infty.$$

(a) Show that $b : \mathbb{R}_+ \to \mathbb{R}$ defined via

$$b(x) := \begin{cases} e^{-\frac{1}{x}} & \text{if } x > 0, \\ 0 & \text{if } x = 0, \end{cases}$$

satisfies the conditions (I) through (V).

Solution:

We have $b(0) = 0$. Since $e^{-\frac{1}{x}}$ is continuous and nondecreasing for $x > 0$ and we have $\lim_{x \downarrow 0} e^{-\frac{1}{x}} = 0$, we conclude that b is continuous and nondecreasing on $\mathbb{R}_+$. We have $e^{-\frac{1}{x}} \leq 1$ for all $x > 0$, so $b(x) \leq 1$ for all $x \in \mathbb{R}_+$, which establishes that b is uniformly bounded above on $\mathbb{R}_+$. Further, b is differentiable on $\mathbb{R}_{++}$, with

$$b'(x) = \frac{1}{x^2} e^{-\frac{1}{x}}, \;\; x > 0,$$

(which is strictly positive for $x > 0$, consistent with b being increasing on $\mathbb{R}_+$). Further,

$$\lim_{x \downarrow 0} b'(x) = \lim_{x \downarrow 0} \frac{1}{x^2} e^{-\frac{1}{x}} = 0,$$

as required. Finally, we have

$$\lim_{x \to \infty} \frac{b(x)}{x b'(x)} = \lim_{x \to \infty} x = \infty.$$

This shows that b satisfies all the conditions required of a.

The point of this part of the question is just to establish that the class of functions a that we are considering is nonempty.

(b) Let $a : \mathbb{R}_+ \to \mathbb{R}$ satisfy the conditions (I) through (V). Define $f : \mathbb{R} \to \mathbb{R}$ with $\mathrm{dom}(f) = \mathbb{R}$ via

$$f(x) := \begin{cases} \int_0^x a(y) dy & \text{if } x \geq 0, \\ f(-x) & \text{if } x < 0. \end{cases}$$

Show that f is a convex twice differentiable function on $\mathbb{R}$.

Solution:

Since a is continuous and uniformly bounded above on $\mathbb{R}$ the function f is well defined. For $x > 0$ we have $f'(x) = a(x)$ and so for $x < 0$ we have $f'(x) = -a(-x)$. Since

$$\lim_{x \downarrow 0} f'(x) = \lim_{x \downarrow 0} a(x) = 0 = \lim_{x \uparrow 0} -a(-x) = \lim_{x \uparrow 0} f'(x),$$

we see that f is also differentiable at 0, with $f'(0) = 0$. We have thus established that f is differentiable on $\mathbb{R}$.

We have $f''(x) = a'(x)$ for $x > 0$ and $f''(x) = a'(-x)$ for $x < 0$. Since we have

$$\lim_{x \downarrow 0} f''(x) = \lim_{x \downarrow 0} a'(x) = 0 = \lim_{x \uparrow 0} a'(-x) = \lim_{x \uparrow 0} f''(x),$$

we see that f is also twice differentiable at 0, with $f''(0) = 0$. We also have $f''(x) \geq 0$ since a is nondecreasing, and so have thus established that f is a convex twice differentiable function on $\mathbb{R}$.

(c) Show that there is $x_0 > 0$ such that the pure Newton method to find the minimum of f, with initial condition x_0, does not converge.

Solution:
Since we have

$$\lim_{x \to \infty} \frac{f'(x)}{x f''(x)} = \lim_{x \to \infty} \frac{a(x)}{x a'(x)} = \infty,$$

we can find $\bar{x} > 0$ such that

$$\frac{f'(x)}{f''(x)} \geq 3x, \quad \text{for all } x > \bar{x}.$$

Pick any $x_0 > \bar{x}$ and consider the pure Newton method initialized at $x^{(0)} = x_0$. Let $(x^{(k)}, k \geq 1)$ denote the sequence of steps of the algorithm. We claim that $|x^{(k)}| \geq 2^k x^{(0)}$. In fact the signs of the $x^{(k)}$ will also alternate between positive and negative.

To see this it suffices to recall that the update equation of the pure Newton method reads

$$x^{(k+1)} = x^{(k)} - \frac{f'(x^{(k)})}{f''(x^{(k)})}.$$

Suppose as an inductive hypothesis that $|x^{(k)}| \geq 2^k x^{(0)}$ (this is certainly true when $k = 0$). Then $|x^{(k)}| > \bar{x}$, so we have $|\frac{f'(x^{(k)})}{f''(x^{(k)})}| \geq 3|x^{(k)}|$ and, further, $\frac{f'(x^{(k)})}{f''(x^{(k)})}$ has the same sign as $x^{(k)}$. We conclude that

$$\begin{aligned}
|x^{(k+1)}| &= \left| x^{(k)} - \frac{f'(x^{(k)})}{f''(x^{(k)})} \right| \\
&\geq \left| \frac{f'(x^{(k)})}{f''(x^{(k)})} \right| - |x^{(k)}| \\
&\geq 2|x^{(k)}|,
\end{aligned}$$

(and in fact $x^{(k+1)}$ has sign opposite to that of $x^{(k)}$). This propagates the inductive step and establishes that the pure Newton method from the given initial condition does not converge. In particular, it does not converge to 0, which is the optimal point for the optimization problem of minimizing f over $\mathbb{R}$.

3 Affine invariance of algorithms

Consider the following unconstrained optimization problem of minimizing a twice differentiable function $f : \mathbb{R}^n \to \mathbb{R}$:

$$\min_{x \in \mathbb{R}^n} f(x). \tag{1}$$

We may make a change of variable transformation $y = Mx$, for an arbitrary but appropriately sized, invertible matrix M and define $g(y) = f(M^{-1}y)$, to obtain the equivalent problem:

$$\min_{y \in \mathbb{R}^n} g(y). \tag{2}$$

If x^* is an optimal solution for (1) then $y^* := Mx^*$ will be an optimal solution for (2).

Consider an algorithm for trying to solve problem (1), which starts at $x^{(0)}$ and updates as $x^{(k)}$ for $k = 1, 2, \ldots$. We may use the same algorithm on problem (2) starting from $y^{(0)}$ to get updates $y^{(k)}$, for $k = 1, 2, \ldots$.

In general, even if we have $y^{(0)} = Mx^{(0)}$, there is no reason to expect that $y^{(k)}$ will equal $Mx^{(k)}$ for $k \geq 1$. If this does happen for all invertible matrices M, all initial conditions $x^{(0)}$, and all $k \geq 1$, we say that the algorithm under consideration is *affine-invariant*.

(a) Show that the pure Newton method is affine-invariant.

Solution:

Let $y^{(0)} := Mx^{(0)}$. For $k \geq 0$, assume as an inductive hypothesis that $y^{(k)} := Mx^{(k)}$, where $(y^{(k)}, k \geq 1)$ denotes the sequence of points generated by the pure Newton method applied to (2) starting from $y^{(0)}$ and $(x^{(k)}, k \geq 1)$ denotes the sequence of points generated by the pure Newton method applied to (1) starting from $x^{(0)}$. The inductive hypothesis certainly holds at $k = 0$.

We can compute

$$\nabla g(y^{(k)}) = M^{-T} \nabla f(M^{-1} y^{(k)}) = M^{-T} \nabla f(x^{(k)}),$$

and

$$\nabla^2 g(y^{(k)}) = M^{-T} \nabla^2 f(M^{-1} y^{(k)}) M^{-1} = M^{-T} \nabla^2 f(x^{(k)}) M^{-1}.$$

Hence we have

$$\begin{aligned} y^{(k+1)} &= y^{(k)} - (\nabla^2 g(y^{(k)}))^{-1} \nabla g(y^{(k)}) \\ &= Mx^{(k)} - (M^{-T} \nabla^2 f(x^{(k)}) M^{-1})^{-1} M^{-T} \nabla f(x^{(k)}) \\ &= Mx^{(k)} - M \nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)}) \\ &= M(x^{(k)} - \nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)})) \\ &= Mx^{(k+1)}. \end{aligned}$$

This propagates the inductive step and thus establishes the affine-invariance of the pure Newton method.

(b) Show that gradient descent with exact line search is not affine-invariant.

Solution:
Let $x \in \mathbb{R}^n$. Gradient descent with exact line search for the problem (1), when at x, will attempt to solve the problem

$$\min_{s \geq 0} f(x - s\nabla f(x)),$$

and will move to the point x^+ that solves this optimization problem. In particular x^+ is on the ray $(x - s\nabla f(x), s \geq 0)$.

Suppose $y = Mx$. Gradient descent with exact line search for the problem (2), when at y, will attempt to solve the problem

$$\min_{s \geq 0} \; g(y - s\nabla g(y)),$$

and will move to the point y^+ that solves this optimization problem. In particular y^+ is on the ray $(y - s\nabla g(y), s \geq 0)$.

For any $s \geq 0$ we have

$$y - s\nabla g(y) = Mx - sM^{-T}\nabla f(x),$$

because $\nabla g(y) = M^{-T}\nabla f(x)$ as we saw in the preceding part of the question. On the other hand, we have

$$M(x - s\nabla f(x)) = Mx - sM\nabla f(x).$$

Thus, unless $M\nabla f(x) = M^{-T}\nabla f(x)$, the ray $(y - s\nabla g(y), s \geq 0)$ will intersect the ray $(M(x - s\nabla f(x)), s \geq 0)$ only when $s = 0$. This means that, in general, we cannot have $y^+ = Mx^+$, so gradient descent with exact line search is not affine-invariant.

As an example, consider $f(x) = x_1^2 + x_2^2$ and

$$M = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}.$$

Then

$$g(y) = f(M^{-1}y) = y_1^2 + \frac{1}{4}y_2^2,$$

so

$$\nabla f(x) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}, \quad \nabla g(y) = \begin{bmatrix} 2y_1 \\ \dfrac{1}{2}y_2 \end{bmatrix}.$$

Also,

$$x^* = 0 = y^*.$$

If $x_0 = [1, 1]^\top$, then $\nabla f(x_0) = [2, 2]^\top$. Since

$$x^* = x_0 - \frac{1}{2}\nabla f(x_0),$$

is already the global minimizer of f and lies on the ray $\{x_0 - s\nabla f(x_0) : s \geq 0\}$, we have $x_1 = x^*$.

On the other hand,

$$y_0 = Mx_0 = \begin{bmatrix} 1 \\ 2 \end{bmatrix},$$

so $\nabla g(y_0) = [2, 1]^\top$. The ray $\{y_0 - s\nabla g(y_0) : s \geq 0\}$ is $\{\begin{bmatrix} 1 - 2s & 2 - s \end{bmatrix}^T : s \geq 0\}$ and so to find y_1 by exact line search we have to solve the problem

$$\min_{s \geq 0} (1 - 2s)^2 + \frac{1}{4}(2 - s)^2.$$

The minimum occurs at $s = \frac{10}{17}$ and gives Then

$$y_1 = \begin{bmatrix} -\frac{3}{17} \\ \frac{24}{17} \end{bmatrix} \neq Mx_1.$$

In fact, the ray $\{y_0 - s\nabla g(y_0) : s \geq 0\}$ does not go through the point $0 = Mx_1$.