EECS 127/227AT **Fall 2020**
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 $x^{(k+1)}$.

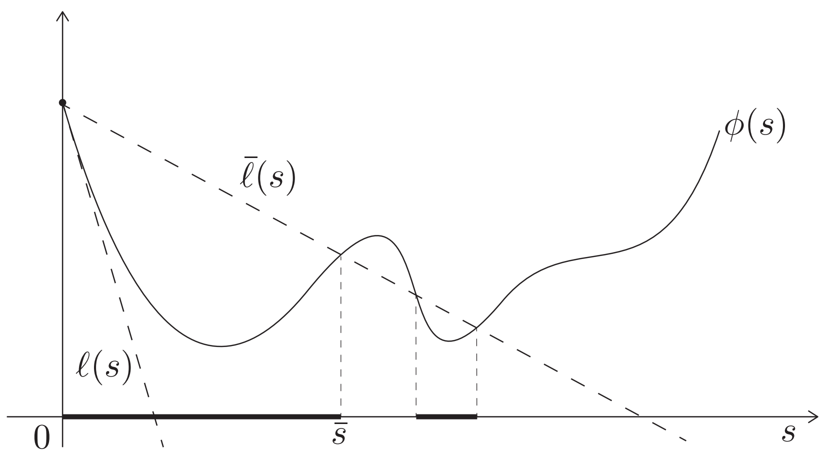

Figure 1: Backtracking line search. Here $l(s) := f(x^{(k)}) - s\|\nabla f(x^{(k)})\|_2^2$ and $\bar{l}(s) := f(x^{(k)}) - \alpha s\|\nabla f(x^{(k)})\|_2^2$. The abscissa is parametrized by s, so the graph is of $\phi(s) := f(x^{(k)} - s\nabla f(x^{(k)}))$. This is only a generic illustration—in this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

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).

(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 $\mathbb{R}$.

(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.

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.

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