Math 408 Homework Set 8 Solutions

(1) Let
$$K := \mathbb{R}^s_- \times \{0\}^{m-s} = \{v \mid 0 \le v_i,\ i = 1, \dots, s,\ v_i = 0\, i = s+1, \dots, m\}.$$

(a) Show that $K^\circ = \mathbb{R}^s_+ \times \mathbb{R}^{m-s}$.

Solution: Let $(w, z) \in \mathbb{R}^s_+ \times \mathbb{R}^{m-s}$. Then, for every $(u, v) \in \mathbb{R}^s_- \times \{0\}^{m-s}$, $\langle (w, z), (u, v)\rangle = \sum_{j=1}^s w_j u_j \le 0$ since $w_j \ge 0$ and $u_j \le 0$, $j = 1, \dots, s$. So $\mathbb{R}^s_+ \times \mathbb{R}^{m-s} \subset K^\circ$. Next let $(w, z) \in K^\circ$. If there is a $j_0 \in \{1, \dots, s\}$ such that $w_{j_0} < 0$, then $\langle (w, z), (-e_{j_0}, 0)\rangle = -w_{j_0} > 0$. But $(-e_{j_0}, 0) \in K$ which contradicts $(w, z) \in K^\circ$. So no such j_0 exists giving $K^\circ \subset \mathbb{R}^s_+ \times \mathbb{R}^{m-s}$.

(b) Given $z \in K$, show that $N_K(z) = \{w \in K^\circ \mid w_i z_i = 0\ i = 1, \dots, s\}$.

Solution: Given $(u, v) \in K$,
$$T_K((u, v)) = \{(p, q) \mid p_i \le 0\ i \in I(u) \text{ and } q = 0\},$$
where $I(u) := \{i \in \{1, \dots, s\} \mid u_i = 0\}$. Therefore,
$$\begin{aligned} N_K((u, v)) &= T_K((u, v))^\circ \\ &= \{(w, z) \mid w_i \ge 0\ i \in I(u),\ w_i = 0\ i \in \{1, \dots, s\} \setminus I(u)\} \\ &= \{(w, z) \in K^\circ \mid w_i u_i = 0\ i = 1, \dots s\}. \end{aligned}$$

(2) Let $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$, and denote the ith row of A by $a_i \in \mathbb{R}^n$, $i = 1, \dots, m$. Let $K := \mathbb{R}^s_- \times \{0\}^{m-s}$ and set $\Omega := \{x \mid Ax - b \in K\}$ and let $\bar{x} \in \Omega$.

(a) Show that
$$T_\Omega(\bar{x}) = \left\{d \in \mathbb{R}^n \ \middle|\ a_i^T d \le 0\ i \in I(\bar{x}) \text{ and } a_i^T d = 0\ i = s+1, \dots, m\right\} = \{d \in \mathbb{R}^n \mid Ad \in T_K(A\bar{x} - b)\} =: A^{-1} T_K(A\bar{x} - b),$$
where $I(\bar{x}) := \left\{i \in \{1, \dots, s\} \ \middle|\ a_i^T \bar{x} = 0\right\}$.

Solution: This follows immediately from the description of the tangent cone to K given in (1)(b) above.

(b) Show that
$$N_\Omega(\bar{x}) = \left\{ \sum_{i\in I(\bar{x})} u_i a_i + \sum_{i=s+1}^m u_i a_i \ \middle|\ 0 \le u_i\ i \in I(\bar{x}) \right\} = \left\{A^T u \mid u \in N_K(Ax - b)\right\} =: A^T N_K(Ax - b).$$

Solution: This follows immediately from the description of the normal cone to K given in (1)(b) above.

(3) Give an example of a smooth nonlinear function $f : \mathbb{R}^n \to \mathbb{R}$, a pair $x, d \in \mathbb{R}^n$, and a number $0 < c < 1$ for which $f'(x; d) < 0$ but the backtracking line-seach fails to terminate in a finite number of steps.

Solution: Let $f(x) := e^{-x} - x$, $x = 0$, $d = 1$, and $c = \frac{1}{2}$. Then $f(x+td) = f(t) = e^{-1} - t < 1 - t = f(x) + ctf'(x; d)$ for all $t > 0$.

(4) Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable with $\nabla F(x)$ invertible on $\mathbb{R}^n$. The Newton iteration for solving $F(x) = 0$ is given by
$$x^{k+1} = x^k - \nabla F(x^k)^{-1} F(x^k).$$
Although this iteration converges at a quadratic rate locally, it may not converge if the initial point x^0 is too far from a solution. To compensate for this deficiency, the Newton iteration is often replaced by a *damped* Newton iteration of the form
$$x^{k+1} = x^k - t_k \nabla F(x^k)^{-1} F(x^k)$$
for some choice of $t_k > 0$. We consider one approach to choosing t_k.

(a) Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ given by $f(x) = \|F(x)\|_2$. Show that if $x \in \mathbb{R}^n$ is such that $F(x) \neq 0$, then f is differentiable at x with

$$\nabla f(x) = \nabla F(x)^T \frac{F(x)}{f(x)} .$$

Solution: Apply the chain rule in conjunction with the fact that the function $\phi(y) = \|y\|_2$ is differentiable at every point except $y = 0$ with $\nabla \phi(y) = y/\|y\|_2$ for $y \neq 0$.

(b) Note that one can attempt to solve $F(x) = 0$ by minimizing the function $f(x)$. This indicates that a damped Newton step size $t_k > 0$ can be computed by doing a line search on the function f in the Newton direction $d_N = -\nabla F(x)^{-1} F(x)$. As a first step, show that the Newton direction d_N is a direction of descent at points x for which $F(x) \neq 0$ by showing that

$$f'(x; d_N) = -\|F(x)\|_2 .$$

Solution: Just plug in to see that

$$f'(x; d_N) = \nabla f(x)^T d_N = -\left(\nabla F(x)^T \frac{F(x)}{f(x)}\right)^T \nabla F(x)^{-1} F(x) = -\|F(x)\|_2 .$$

(c) Show that the backtracking line search applies and takes the form

$$\begin{array}{rll} t_k = & \max & \gamma^s \\ & \text{s.t.} & s \in \{0, 1, \dots\} \\ & & \|F(x^k + \gamma^s d_N^k)\|_2 \leq (1 - c\gamma^s)\|F(x^k)\|_2 \end{array} ,$$

where $\gamma \in (0, 1)$ and $c \in (0, 1)$ are the line search parameters.

Solution: We have

$$f(x + td_N) \leq f(x) + ctf'(x; d_N) \quad \iff \quad \|F(x + td_N)\|_2 \leq \|F(x)\|_2 - ct\|F(x)\|_2 = (1 - ct)\|F(x)\|_2.$$

(5) Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be continuously differentiable and consider the problem of finding a point x such that $F(x) = 0$. This is the same problem that led to the development of Newton's method except now we are interested in the case where $n \neq m$. If $m < n$, then there are most likely infinitely many solutions. On the other-hand, if $m > n$, then there are probably no solutions. In this problem we consider the case when $m > n$. Since it is unlikely that there exists an x satisfying $F(x) = 0$, we instead try to find an x that makes $\|F(x)\|$ as small as possible. For this we define $f(x) = \frac{1}{2}\|F(x)\|_2^2$. It is easily shown (try it) that

$$\begin{aligned} \nabla f(x) &= \nabla F(x)^T F(x) \\ \nabla^2 f(x) &= \nabla F(x)^T \nabla F(x) + \sum_{i=1}^m F_i(x) \nabla^2 F_i(x) , \end{aligned}$$

where the functions $F_i : \mathbb{R}^n \to \mathbb{R}$ are the component functions of F.

(a) In this setting we still try to follow the path suggested by Newton where we successively solve for the search direction using the linearization of F. But in this case the search direction is given as the solution to the linear least squares problem

$$\mathcal{GN} \qquad \min_{d \in \mathbb{R}^n} \frac{1}{2}\|F(x^k) + \nabla F(x^k) d\|_2^2 .$$

Methods using the solution to this problem as a search direction are called *Gauss-Newton* methods. Denote a solution by d_{GN}. Under what conditions does this problem have a unique solution?

Solution: Our work on the linear least-squares problem in Chapter 1 tells us that this problem has a unique solution if and only if $\text{Nul}\left(\nabla F(x^k)\right) = \{0\}$.

(b) When a unique solution to the problem $\mathcal{GN}$ exists give a formula for d_{GN}.

Solution: Again from Chapter 1, $d_{GN} = -(\nabla F(x^k)^T \nabla F(x^k))^{-1} \nabla F(x^k)^T F(x^k)$.

(c) Use this formula for d_{GN} to show that it is a descent direction for f giving a formula for $f'(x^k; d_{GN})$.

Solution: We Have

$$f'(x^k; d_{GN}) = \nabla f(x^k)^T d_{GN} = -(\nabla F(x^k)^T F(x^k))^T (\nabla F(x^k)^T \nabla F(x^k))^{-1} \nabla F(x^k)^T F(x^k) < 0$$

whenever $\nabla F(x^k)^T F(x^k) \neq 0$ since $(\nabla F(x^k)^T \nabla F(x^k))^{-1}$ is positive definite when $\mathrm{Nul}\left(\nabla F(x^k)\right) = \{0\}$.

(d) Show that d_{GN} is also a descent direction for the function $h(x^k) = \|F(x^k)\|_2$.

Solution:

$$h'(x^k; d_{GN}) = -\left(\nabla F(x^k)^T \frac{F(x^k)}{\|F(x^k)\|_2}\right)^T (\nabla F(x^k)^T \nabla F(x^k))^{-1} \nabla F(x^k)^T F(x^k) = \frac{f'(x^k; d_{GN})}{\|F(x^k)\|_2} < 0.$$

(e) Assuming that $\nabla F(x)$ is continuous, show that

$$h'(x; d) \leq \|F(x) + \nabla F(x) d\|_2 - h(x) \ .$$

Solution: First observe that

$$h(y) = \|F(x) + \nabla F(x)(y - x) + o(\|y - x\|_2)\|_2 = \|F(x) + \nabla F(x)(y - x)\|_2 + o(\|y - x\|_2),$$

so that

$$h'(x; d) = \lim_{t \downarrow 0} \frac{\|F(x) + t\nabla F(x) d\|_2 - \|F(x)\|_2}{t}.$$

Next note that for $0 < t < 1$, the convexity of the norm implies that

$$\begin{aligned} \|F(x) + t\nabla F(x)d\|_2 - \|F(x)\|_2 &= \|(1-t)F(x) + t(F(x) + \nabla F(x)d)\|_2 - \|F(x)\|_2 \\ &\leq (1-t)\|F(x)\|_2 + t\|F(x) + \nabla F(x)d\|_2 - \|F(x)\|_2 \\ &= t\left(\|F(x) + \nabla F(x)d\|_2 - \|F(x)\|_2\right). \end{aligned}$$

Combining this inequality with the expression for $h'(x; d)$ given above yields the result.