Problem Set 3 Solutions

March 3, 2005

1.3.4

Without loss of generality we assume that $x^* = 0$, so the iteration is written as

$$x^{k+1} = x^k - s(Qx^k + e^k) = (I - sQ)x^k - se^k.$$

Thus, we have

$$\|x^{k+1}\| \leq \|(I - sQ)x^k\| + s\|e^k\| \leq q\|x^k\| + s\delta.$$

Applying sequentially this inequality, we obtain

$$\|x^k\| \leq q^k\|x^0\| + s\delta(1 + q + \cdots + q^{k-1}),$$

from which

$$\|x^k\| \leq q^k\|x^0\| + \frac{s\delta}{1-q}.$$

1.4.1

Consider the problem

$$\min_y h(y) = f(Sy).$$

Newton's method generates a sequence according to

$$y^{k+1} = y^k - \alpha^k \left(\nabla^2 h(y^k)\right)^{-1} \nabla h(y^k).$$

We have

$$\nabla h(y) = S'\nabla f(Sy),$$

$$\nabla^2 h(y) = S'\nabla^2 f(Sy)S.$$

So Newton's method in the space of y can be re-written as

$$Sy^{k+1} = Sy^k - \alpha^k S \left(\nabla^2 h(y^k)\right)^{-1} \nabla h(y^k)$$

$$= Sy^k - \alpha^k S \left(S'\nabla^2 f(Sy^k)S\right)^{-1} S'\nabla f(Sy^k)$$

$$= Sy^k - \alpha^k SS^{-1} \left(\nabla^2 f(Sy^k)\right)^{-1} (S')^{-1}S'\nabla f(Sy^k)$$

$$= Sy^k - \alpha^k \left(\nabla^2 f(Sy^k)\right)^{-1} \nabla f(Sy^k)$$

By replacing Sy^k with x^k, we have

$$x^{k+1} = x^k - \alpha^k \left(\nabla^2 f(x^k)\right)^{-1} \nabla f(x^k),$$

which is Newton's method in the space of the variables x.

1.4.8

(a) We have

$$\nabla f(x) = \beta \|x\|^{\beta-2} x,$$
$$\nabla^2 f(x) = \beta(\beta-2)\|x\|^{\beta-4} xx' + \beta\|x\|^{\beta-2} I.$$

We guess that the Newton direction has the form $d = -\gamma x$, where γ is a scalar, and we check the equation $\nabla^2 f(x) d = -\nabla f(x)$ to determine the appropriate value of γ. In this way, we obtain

$$\gamma = -\frac{1}{\beta-1}.$$

Alternatively, we could use the formula

$$(A + CBC')^{-1} = A^{-1} - A^{-1}C(B^{-1} + C'A^{-1}C)^{-1}C'A^{-1}$$

[Eq. (A.11) from Appendix A] to find the inverse of $\nabla^2 f(x)$. Thus Newton's method has the form

$$x^{k+1} = x^k - \frac{1}{\beta-1} x^k = \frac{\beta-2}{\beta-1} x^k.$$

Therefore

$$||x^{k+1}|| = \left|\frac{\beta-2}{\beta-1}\right| ||x^k||.$$

Hence, if $\beta > 3/2$ then $\left|\frac{\beta-2}{\beta-1}\right| < 1$ and the method converges to 0 for any initial point x^0. In particular, for $\beta = 2$ it converges in one step as expected. If $\beta = 3/2$, the method generates the points x^k on the sphere $S = \{x \mid ||x|| = ||x^0||\}$ and does not converge for any initial point $x^0 \neq 0$. For $1 < \beta < 3/2$, we have $\left|\frac{\beta-2}{\beta-1}\right| > 1$ and the method diverges for all $x^0 \neq 0$.

Now let's look at the case when $\beta \leq 1$. Since

$$(\nabla^2 f(x))^{-1} = \frac{1}{\beta||x||^{\beta-2}} \left(I - \frac{\beta-2}{\beta-1} \cdot \frac{xx'}{||x||^2}\right),$$

we see that $(\nabla^2 f(x))^{-1}$ does not exist for $\beta = 1$. If $\beta < 1$, then $\left|\frac{\beta-2}{\beta-1}\right| = \frac{2-\beta}{1-\beta} > 1$ and the method diverges for any initial point.

(b) With the Armijo rule we have

$$\|x^{k+1}\| = \left|1 - \frac{\alpha^k}{\beta-1}\right| \|x^k\|,$$

At each step, Armijo rule sets the stepsize $\alpha^k = a^m s$, where s is the initial stepsize, a is the reduction factor, and m is the smallest nonnegative integer for which

$$f(x^{k+1}) - f(x^k) \le \sigma a^m s \nabla f(x^k)' d^k.$$

For this problem, this test becomes

$$\|x^k\|^\beta \left(1 - \left|1 - \frac{a^m s}{\beta - 1}\right|^\beta\right) \ge \sigma a^m s \frac{\beta \|x^k\|^\beta}{\beta - 1},$$

which implies that the stepsize is the same at each iteration, i.e., $\alpha^k = \alpha$ for all k. Hence we have

$$\|x^{k+1}\| = \left|1 - \frac{\alpha}{\beta - 1}\right| \|x^k\|,$$

and also $\left|1 - \frac{\alpha}{\beta-1}\right| < 1$, therefore the method converges for any starting point when $\beta > 1$.

1.5.1

(a) We have

$$f(x) = \frac{1}{2}||g(x)||^2 = \frac{1}{2}\sum_{i=1}^{m} ||g_i(x)||^2.$$

The Hessian matrix is given by

$$\nabla^2 f(x^*) = \nabla g(x^*) \nabla g(x^*)' + \sum_{i=1}^{m} \nabla^2 g_i(x^*) g_i(x^*).$$

At any optimal solution x^* for which $g(x^*) = 0$, we have that $g_i(x^*) = 0$ for $1 \le i \le m$. Hence, at such points the Hessian has the following form

$$\nabla^2 f(x^*) = \nabla g(x^*) \nabla g(x^*)'.$$

Note that $g(x^*)$ is an $n \times m$ matrix. Then $Ra(\nabla g(x^*)')$ has dimension at most m, where $Ra(A)$ denotes the range of the matrix A (the linear space spanned by the columns of A). Consequently, the dimension of the null space of A is at least $n - m > 0$. Hence, there is a vector $v \neq 0$ for which $\nabla g(x^*)' v = 0$. This implies $\nabla^2 f(x^*) v = 0$, i.e. $\nabla^2 f(x^*)$ is singular.

(b) The function $f(x) = \frac{1}{2}||z - Ax||^2$ attains a minimum at x^*, since it is positive semidefinite and bounded from below (see Section 1.1). Since $Ra(A)$ has dimension at most m, the null space of A has dimension at least $n - m > 0$. Therefore, there exists a nonzero vector v for which $Av = 0$. Let x^* be a solution to the problem. Then, for any scalar λ the vector $x^* + \lambda v$ is also solution, since

$$f(x^* + \lambda v) = \frac{1}{2}||z - Ax^* - \lambda A v||^2 = ||z - Ax^*||^2 = f(x^*).$$

Thus, there are infinitely many optimal solutions.

When A has linearly independent rows, then AA' is an invertible $m \times m$ matrix so that the point $x^* = A'(AA')^{-1} z$ is well defined. For this point we have

$$\nabla f(x^*) = A'Ax^* - A'z = A'AA'(AA')^{-1} z - A'z = 0.$$

Since $\nabla^2 f(x) = A'A \geq 0$ for all x, the first order necessary optimality conditions are also sufficient, i.e. $x^* = A'(AA')z$ is one of the solutions for the given problem.

1.5.5

We have

$$x^{k+1} - x^* = x^k - \alpha Q_k x^k - x^* = (I - \alpha Q_k)(x^k - x^*) - \alpha \nabla f_k(x^*),$$

since $Q_k x^* = \nabla f_k(x^*)$. We then have

$$\begin{aligned} \|x^{k+1} - x^*\| &\leq \|(I - \alpha Q_k)(x^k - x^*)\| + \alpha \|\nabla f_k(x^*)\| \\ &\leq \|(I - \alpha Q_k)(x^k - x^*)\| + \alpha\epsilon \\ &\leq \|I - \alpha Q_k\| \|x^k - x^*\| + \alpha\epsilon. \end{aligned}$$

Note that $\|I - \alpha Q_k\| = \max\{|1 - \alpha\gamma|, |1 - \alpha\Gamma|\}$. Analyzing these terms individually, we have

$$1 - \alpha\gamma \geq 1 - \frac{2\gamma}{\gamma + \Gamma} = \frac{\Gamma - \gamma}{\Gamma + \gamma} \geq 0,$$

so $|1 - \alpha\gamma| = 1 - \alpha\gamma$. On the other hand,

$$1 - \alpha\Gamma \geq \frac{\gamma - \Gamma}{\gamma + \Gamma},$$

and combining these facts, we have that $\|I - \alpha Q_k\| = 1 - \alpha\gamma$. We now distinguish the two cases.

Case 1: $\|x^k - x^*\| > 2\epsilon/\gamma \Rightarrow \alpha\epsilon < \alpha\gamma/2\|x^k - x^*\|$. So we have

$$\begin{aligned} \|x^{k+1} - x^*\| &< (1 - \alpha\gamma)\|x^k - x^*\| + \frac{\alpha\gamma}{2}\|x^k - x^*\| \\ &= (1 - \frac{\alpha\gamma}{2})\|x^k - x^*\|. \end{aligned}$$

Case 2: $\|x^k - x^*\| \leq 2\epsilon/\gamma$. This gives us

$$\begin{aligned} \|x^{k+1} - x^*\| &\leq (1 - \alpha\gamma)\frac{2\epsilon}{\gamma} + \alpha\epsilon \\ &= \frac{2\epsilon}{\gamma} - \alpha\epsilon \leq \frac{2\epsilon}{\gamma}. \end{aligned}$$

The desired result now follows: $\|x^k - x^*\|$ decreases geometrically until we have $\|x^k - x^*\| \leq 2\epsilon/\gamma$, after which we are guaranteed to remain in this "close" region.