Recitation 1 Solutions

February 20, 2005

1.1.9

For any $x, y \in \mathcal{R}^n$, from the second order expansion (see Appendix A, Proposition A.23) we have

$$f(y) - f(x) = (y-x)'\nabla f(x) + \frac{1}{2}(y-x)'\nabla^2 f(z)(y-x), \qquad (1)$$

where z is some point of the line segment joining x and y. Setting $x = 0$ in (1) and using the given property of f, it can be seen that f is coercive. Therefore, there exists $x^* \in \mathcal{R}^n$ such that $f(x^*) = \inf_{x\in\mathcal{R}^n} f(x)$ (see Proposition A.8 in Appendix A). The condition

$$m||y||^2 \leq y'\nabla^2 f(x)y, \qquad \forall\, x, y \in \mathcal{R}^n,$$

is equivalent to strict convexity of f. Strict convexity guarantees that there is a unique global minimum x^*. By using the given property of f and the expansion (1), we obtain

$$(y-x)'\nabla f(x) + \frac{m}{2}||y-x||^2 \leq f(y) - f(x) \leq (y-x)'\nabla f(x) + \frac{M}{2}||y-x||^2.$$

Taking the minimum over $y \in \mathcal{R}^n$ in the expression above gives

$$\min_{y\in\mathcal{R}^n}\left((y-x)'\nabla f(x) + \frac{m}{2}||y-x||^2\right) \leq f(x^*) - f(x) \leq \min_{y\in\mathcal{R}^n}\left((y-x)'\nabla f(x) + \frac{M}{2}||y-x||^2\right).$$

Note that for any $a > 0$

$$\min_{y\in\mathcal{R}^n}\left((y-x)'\nabla f(x) + \frac{a}{2}||y-x||^2\right) = -\frac{1}{2a}||\nabla f(x)||^2,$$

and the minimum is attained for $y = x - \frac{\nabla f(x)}{a}$. Using this relation for $a = m$ and $a = M$, we obtain

$$-\frac{1}{2m}||\nabla f(x)||^2 \leq f(x^*) - f(x) \leq -\frac{1}{2M}||\nabla f(x)||^2.$$

The first chain of inequalities follows from here. To show the second relation, use the expansion (1) at the point $x = x^*$, and note that $\nabla f(x^*) = 0$, so that

$$f(y) - f(x^*) = \frac{1}{2}(y-x^*)'\nabla^2 f(z)(y-x^*).$$

The rest follows immediately from here and the given property of the function f.

1.2.10

We have

$$\nabla f(x) - \nabla f(x^*) = \int_0^1 \nabla^2 f(x^* + t(x - x^*))(x - x^*)dt$$

and since

$$\nabla f(x^*) = 0,$$

we obtain

$$(x - x^*)'\nabla f(x) = \int_0^1 (x - x^*)'\nabla^2 f(x^* + t(x - x^*))(x - x^*)dt \geq m \int_0^1 \|x - x^*\|^2 dt.$$

Using the Cauchy-Schwartz inequality $(x - x^*)'\nabla f(x) \leq \|x - x^*\| \, \|\nabla f(x)\|$, we have

$$m \int_0^1 \|x - x^*\|^2 dt \leq \|x - x^*\| \, \|\nabla f(x)\|,$$

and

$$\|x - x^*\| \leq \frac{\|\nabla f(x)\|}{m}.$$

Now define for all scalars t,

$$F(t) = f(x^* + t(x - x^*))$$

We have

$$F'(t) = (x - x^*)'\nabla f(x^* + t(x - x^*))$$

and

$$F''(t) = (x - x^*)'\nabla^2 f(x^* + t(x - x^*))(x - x^*) \geq m\|x - x^*\|^2 \geq 0.$$

Thus F' is an increasing function, and $F'(1) \geq F'(t)$ for all $t \in [0, 1]$. Hence

$$f(x) - f(x^*) = F(1) - F(0) = \int_0^1 F'(t)dt \leq F'(1) = (x - x^*)'\nabla f(x) \leq \|x - x^*\| \, \|\nabla f(x)\| \leq \frac{\|\nabla f(x)\|^2}{m},$$

where in the last step we used the result shown earlier.

1.2.8

By the definition of d^k, we have

$$||d^k|| = \max_{1 \leq i \leq n} |\frac{\partial f(x^k)}{\partial x_i}| \leq ||\nabla f(x^k)||.$$

If $\{x^k\}$ is a sequence converging to some x with $\nabla f(\bar{x}) \neq 0$, then $\nabla f(x_k) \to \nabla f(\bar{x})$, so that $\{\nabla f(x^k)\}$ is bounded, which in view of the above relation implies that $\{d^k\}$ is bounded. Furthermore, since

$$\nabla f(x^k)'d^k = -\max_{1 \leq i \leq n} |\frac{\partial f(x^k)}{\partial x_i}|^2 \leq -\frac{1}{n} \sum_{i=1}^{n} |\frac{\partial f(x^k)}{\partial x_i}|^2 = -\frac{1}{n}||f(x^k)||^2,$$

it follows that

$$\limsup_{k \to \infty} \nabla f(x^k)'d^k \leq -\frac{1}{n} \lim_{k \to \infty} ||\nabla f(x^k)||^2 = -\frac{1}{n}||\nabla f(\bar{x})||^2 < 0.$$

Therefore the sequence $\{d^k\}$ is gradient related, and the result follows from Proposition 1.2.1.