APM462H1S: Nonlinear optimization, Winter 2014.

Summary of January 29 lecture.

The January 29 lecture was almost entirely focused on the *method of steepest descent*, described in section 8.6 of the textbook.

IMPORTANT: See the correction on page 2 below concerning the definition of the condition number of a symmetric matrix.

In particular we discussed the following topics:

1. A proof that the method of steepest descent converges.

This was done by verifying that it satisfies the hypotheses of the "Global convergence theorem" from Section 7.7 of the textbook (Luenberger and Ye), which was discussed in the January 20 lecture.

The proof of the convergence of the method of steepest descent can be found in the textbook, partly in section 8.4 and partly in the beginning of section 8.6. I combined these into a single proof and tried to simplify it in some minor ways, but the idea of the discussion in class is the same as in the book.

2. detailed consideration of an explicit example of the method of steepest descent, for minimizing the fuction

$$f(x,y) = \frac{1}{2}x^2 + \frac{5}{0}y^2.$$

We found that if we start at $(x_0, y_0) = (100, 1)$ then

$$\begin{aligned}
(x_1, y_1) &= \frac{99}{101}(100, -1) \\
(x_2, y_2) &= (\frac{99}{101})^2(100, 1) \\
(x_3, y_3) &= (\frac{99}{101})^3(100, -1)
\end{aligned}$$

and generally,

$$(x_k, y_y) = (\frac{99}{101})^k(100, (-1)^k).$$

Thus, for the particular initial guess that we considered, the method converges pretty slowly.

Although we did not say it, from the above conclusions it is easy to see that

$$f(x_k, y_k) = (\frac{99}{101})^{2k} f(x_0, y_0)$$

This also converges quite slowly.

3. We discussed general convergence properties of the method of steepest descent for a quadratic function

$$f(x) = \frac{1}{2}x^TQx - b^Tx.$$

We defined

$$x^* = Q^{-1}b = \text{ minimizer of } f$$

and

$$E(x) = \frac{1}{2}(x - x^*)^TQ(x - x^*).$$

We also checked that $E(x) = f(x) - \frac{1}{2}(x^*)^TQx^* = f(x) - f(x^*)$, so E and f have the same minimum points.

Under these conditions, the main fact is that if the sequence $x_0, x_1, \ldots$ is generated by the method of steepest descent, then

$$\boxed{E(x_{k+1}) \leq (\frac{r-1}{r+1})^2 E(x_k), \qquad \text{where } r \text{ is the condition number of } Q,} \tag{1}$$

and the condition number is defined by:

$$\boxed{r = \text{ condition number of } Q \; = \; \frac{\text{largest eigenvalue of } Q}{\text{smallest eigenvalue of } Q} \; \geq 1.}$$

IMPORTANT!! In the lecture I gave the ***wrong definition*** of the condition number (what I defined was the *reciprocal* of the condition number). The definition given here is the correct one. As a result of this correction, some other formulas now look a little different as well. **I will try to minimize the confusion this error might cause,** for example by reminding you of the correct formulas at every opportunity.

We want $E(x_{k+1})$ to decrease to zero as quickly as possible, so the worst case is when $E(x_{k+1})$ is as large as possible, ie $E(x_{k+1}) = (\frac{r-1}{r+1})^2 E(x_k)$, and $\frac{r-1}{r+1}$ is close to 1. This can actually happen. Indeed, it happens in the explicit example discussed above, since the function $f(x, y) = \frac{1}{2}x^2 + 50y^2$ can be written

$$f = \frac{1}{2}x^TQx \qquad \text{where } Q = \begin{pmatrix} 1 & 0 \\ 0 & 100 \end{pmatrix}$$

with condition number $r = 100$ (large) and $\frac{r-1}{r+1} = \frac{99}{101}$ (close to 1).

In general, we see that

$$\text{the method of steepest descent} \begin{cases} \text{converges well} & \text{if } r \approx 1 \\ \text{may converge badly} & \text{if } r \gg 1 \end{cases}$$

4. We went over *Application 1* and *Application 2* from Section 8.7 of the textbook. Some points related to *Application 2* will also appear on the next homework asignment.

5. We went over the proof of the key fact (1). That material is a little too hard for this course, and you should not worry much about it – we are mainly interested in knowing when we need to be careful about bad convergence for steepest descent. But if you want to consult the proof, some details that I skipped over can be found in section 8.6 of the textbook.

6. Finally, we quickly mentioned the following fact:

Assume that f is a function such that[1]

$$aI \le \nabla^2 f(x) \le AI \qquad \text{for all } x,$$

where I is the identity matrix and $0 < a \le A$.

Next, define $E(x) = f(x) - f(x^*)$, where x^* is the global minimum point.

Then a sequence $x_0, x_1, x_2, \dots$ generated the method of steepest descent satisfies

$$\boxed{E(x_{k+1}) \le (1 - \frac{a}{A})E(x_k).}$$

This is proved in Section 8.6 of the textbook. We will not use the proof, and for us the main point is that, although we derived the notion of *condition number* while considering quadratic minimization problems, is still relevant for more general convex minimization problems. In particular, for every x, the smallest eigenvalue of $\nabla^2 f(x)$ is at least a, and the largest eigenvalue of $\nabla^2 f(x)$ is at most A, so

$$\frac{A}{a} \ge \sup_x \text{ (condition number of } \nabla^2 f(x)).$$

[1]The notation means the following: if Q and R are symmetric, $n \times n$ matrices, then $Q \le R$ means that

$$v^T Q v \le v^T R v \text{ for all vectors } v \in E^n.$$

Also, note that when s is a number and I is the identity matrix, $v^T(sI)v = s(v^T I v) = s v^T v = s|v|^2$. So

$$aI \le Q \le AI, \qquad \text{if and only if} \qquad a|v|^2 \le v^T Q v \le A|v|^2 \text{ for all } v \in E^n.$$

It is also not hard to check that

$$aI \le Q \le AI, \qquad \text{if and only if} \qquad a \le \lambda \le A, \text{ for every eigenvalues } \lambda \text{ of } Q.$$