APM462H1S: Nonlinear optimization, Winter 2014.

Summary of February 10 and 24 lectures.

The lectures on February 10 and 24 covered material from Chapter 11 of the textbook, sections 1,2,3,5,6, and 8, including

- a rather complete discussion of sections 1,2,3 and 5
- a brief discussion of an example, related somewhat to section 6; and
- the first-order necessary conditions from section 8.

The rest of these notes discuss topics from Section 6 of Chapter 11, filling in some details that were skipped over rather quickly in the lecture.

an example. When discussing second-order conditions for problems with equality constraints: we considered the example problem of minimizing

$$f(x_1, x_2, x_3, x_4) = -[x_1x_2x_3 + x_1x_2x_4 + x_1x_3x_4 + x_2x_3x_4]$$

subject to the constraint

$$h(x_1, \ldots, x_4) = 0, \qquad \text{for} \quad h(x_1, \ldots, x_4) = x_1 + x_2 + x_3 + x_4 - 4.$$

We first compute

$$\nabla f = [x_2x_3+x_2x_4+x_3x_4,\ x_1x_3+x_1x_4+x_3x_4,\ x_1x_2+x_1x_4+x_2x_4,\ x_1x_2+x_1x_3+x_2x_3]$$

Thus the first-order conditions

$$\nabla f + \lambda \nabla h = 0$$

can be written out as

$$\begin{aligned}
-(x_2x_3 + x_2x_4 + x_3x_4) + \lambda &= 0\\
-(x_1x_3 + x_1x_4 + x_3x_4) + \lambda &= 0\\
-(x_1x_2 + x_1x_4 + x_2x_4) + \lambda &= 0\\
-(x_1x_2 + x_1x_3 + x_2x_3) + \lambda &= 0.
\end{aligned}$$

Combined with the constraint equation $h = 0$, this gives 5 equations for the 5 unknowns $x_1, \ldots, x_4, \lambda$.

Since the equations are so symmetric with respect to $x_1, \ldots, x_4$, we might guess that there should be a solution with $x_1 = x_2 = x_3 = x_4$. Having guessed this, we can then easily verify that in fact a solution is

$$x_1 = x_2 = x_3 = x_4 = 1, \quad \lambda = -3.$$

Now we want to use the second-order conditions to check whether this is a local minimum. So we consider the matrix

$$L = \nabla^2 f - \lambda \nabla^2 h = -\begin{pmatrix} 0 & x_3 + x_4 & x_2 + x_4 & x_2 + x_3\\ x_3 + x_4 & 0 & x_1 + x_4 & x_1 + x_3\\ x_2 + x_4 & x_1 + x_4 & 0 & x_1 + x_2\\ x_2 + x_3 & x_1 + x_3 & x_1 + x_2 & 0 \end{pmatrix}$$

At the point $(1, 1, 1, 1)$ this reduces to

$$L = -\begin{pmatrix} 0 & 2 & 2 & 2\\ 2 & 0 & 2 & 2\\ 2 & 2 & 0 & 2\\ 2 & 2 & 2 & 0 \end{pmatrix}$$

To examine the second-order conditions, we need to check whether

$$y^T L y > 0 \quad \text{ for all } y \text{ such that } \nabla h(x^*) y = 0.$$

or equivalently

$$y^T L y > 0 \quad \text{ for all } y \text{ such that } y_1 + y_2 + y_3 + y_3 = 0.$$

First solution One quick but sneaky way to do this is to note that

$$L = 2\left[\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \end{pmatrix} \right] = 2(\,\mathrm{I} - \nabla h^T \nabla h)$$

Here of course I denotes the identity matrix. Recall also that, following the textbook, we always think of the gradient as a row vector, so $\nabla h^T \nabla h$ is a $n \times n$ matrix. Thus if $\nabla h\ y = 0$,

$$y^T L y = 2(y^T \mathrm{Id}\, y - y^T \nabla h^T \nabla h\ y) = 2 y^T y = 2|y|^2,$$

where we have used the assumption that $\nabla h\ y = 0$ as well as the obvious fact that $\mathrm{I}\, y = y$ for any y. This says that L is positive definite in directions orthogonal to ∇h, which implies that f has a strict local minimum at $x^* = (1,1,1,1)$.

Second solution. If we want to solve this in a more systematic way, we can proceed as follows:

step 1. Pick a basis for the vector space

$$M = \{y \in \mathbb{R}^4 : \nabla h\ y = 0\} = \{y \in \mathbb{R}^4 : y_1 + y_2 + y_3 + y_4 = 0\}.$$

It can be an orthonormal basis but it does not have to be. For example, we can choose

$$v_1 = (1, -1, 0, 0), \qquad v_2 = (1, 0, -1, 0), \qquad v_3 = (1, 0, 0, -1).$$

We know this is a basis because these three vectors are linearly independent (this is easy to see) and M is a 3-dimensional vector space.

step 2. consider the 2×2 matrix whose (i,j) entry is $v_i^T L v_j$. Let's call this matrix L_M. For the basis we have chosen above, it is easy to see that in fact $L v_i = 2 v_i$ for $i = 1,2,3$, and then it is straightforward to check that

$$L_M = \begin{pmatrix} 8 & 4 & 4 \\ 4 & 8 & 4 \\ 4 & 4 & 8 \end{pmatrix}.$$

step 3. Check whether this matrix is positive definite. We know that this holds if and only if

$$\det(8) > 0, \qquad \det \begin{pmatrix} 8 & 4 \\ 4 & 8 \end{pmatrix} > 0, \qquad \text{and} \qquad \det \begin{pmatrix} 8 & 4 & 4 \\ 4 & 8 & 4 \\ 4 & 4 & 8 \end{pmatrix} > 0.$$

(These are the $1 \times 1, 2 \times 2$ and 3×3 matrices in the upper left corner of L_M, so to speak.) This is in fact true, as can be checked in a few minutes of calculations. So we reach the same conclusion as before: the point x^* is a local minimum of f subject to the constraint h.

a variant of the second solution. It would of course be possible to follow the same procedure as in the second solution, but with a different choice of the vectors

v_1, v_2, v_3. Certain choices would lead to matrices L_M that may be easier to work with.

For example, in the above problem we can choose orthogonal vectors such as

$$v_1 = (1, -1, 0, 0), \qquad v_2 = (1, 1, -2, 0), \qquad v_3 = (1, 1, 1, -3).$$

Then it turns out again that $Lv_i = 2v_i$ for every i, and hence one can check that

$$L_M = \begin{pmatrix} 8 & 0 & 0 \\ 0 & 24 & 0 \\ 0 & 0 & 48 \end{pmatrix}.$$

which is clearly positive definite.

a remark : in the above second solution (and its variant), once we note that the basis vectors v_1, v_2 and v_3 all satisfy $Lv_i = 2v_i$, we can conclude that $Lv - 2v$ for every $v \in M$. This is true because every $y \in M$ can be written in the form $y = a_1 v_1 + a_2 v_2 + a_3 v_3$ for some coefficients a_1, a_2, a_3, so that

$$Ly = a_1 Lv_1 + a_2 Lv_2 + a_3 Lv_3 = 2a_1 v_1 + 2a_2 v_2 + 2a_3 v_3 = 2y.$$

It follows that $y^T Ly = 2|y|^2$, as we found in our first ("quick but sneaky") solution.