APM462H1S, Winter 2014 , About the Midterm

The midterm for APM462 will take place on Monday, March 3 from 6:10 - 8pm in BA1170.

It will cover the following material:

(1) minimization problems in E^n (sections 7.1 through 7.3)
   (a) first- and second-order necessary conditions for a local minimum
   (b) sufficient condition for a local minimum

(2) minimization problems in a subset Ω of E^n
   (a) first-order necessary condition for a local minimum.

(3) convex functions (see Section 7.4-5 of the textbook) including
   (a) definition of convexity.
   (b) a C^1 function f is convex if any only if $f(y) \geq f(x) + \nabla f(x)(y - x)$ for all x and y.
   (c) a C^2 function f is convex if and only if its Hessian matrix $\nabla^2 f(x)$ is positive semidefinite at every x.
   (d) Every critical point of a convex function is a global minimum point.

(4) the Global Convergence Theorem for iterative methods for solving minimization problems

   (from Section 7.7 of the textbook or – just as good – the slightly simpler statement presented in the lecture, see below. For this you have to understand things like the definition of a *closed point-to-set mapping.*)

(5) iterative methods for minimizing functions of a single variable (section 8.2)
   (a) Newton's method
   (b) more generally, the idea of "curve fitting" methods: given a function f to be minimized, and a point x_i, find x_{i+1} by approximating f near x_i by a simpler function (eg quadratic) and minimizing the simpler function. See Section 8.2 of the textbook.

(6) the method of steepest descent (section 8.6 and parts of 8.4)
   (a) the definition of the method of steepest descent.
   (b) the condition number of a matrix, and how it determines convergence properties (for quadratic problems).
   It would be a good idea to memorize the formula $E(x_{n+1}) \leq (\frac{r-1}{r+1})^2 E(x_n)$ where $E(x) = f(x) - f(x^*)$, and f is a quadratic function, and r is the condition number of Q (the matrix of second derivatives of f.)

(7) conjugate directions and conjugate gradient methods (sections 9.1 - 9.3)
   (a) definition of Q-orthogonal vectors
   (b) definition of the conjugate directions method
   (c) the idea of the conjugate gradient method: like the conjugate directions method, except that the directions $d_1, d_2, \dots$ are determined iteratively: d_{k+1} is found by taking $g_{k+1} = \nabla f(x_{k+1})^T$ and "correcting it" to make it Q-orthogonal to d_k.
   (d) convergence properties of conjugate directions methods (including the conjugate gradient method.)

You also should know linear algebra and calculus at the level of this class (see the notes on Blackboard), especially including topics that have appeared on the first two homework assignments.

The version of the Global Convergence Theorem (section 7.7 of the textbook) that we presented in the lecture was slightly simpler than the one in the book, and went like this:

Global Convergence Theorem *Assume that X is either E^n or a closed subset Ω of E^n, and that we want to find a global minimum of a function f defined on X.*

Let Γ be the set of global minimum points of f in X.

Let A be a point-to-set mapping on X, satisfying

(1) *A is closed at x, for every $x \notin \Gamma$.*

(2) *if $y \in A(x)$, then $f(y) \leq f(x)$, with strict inequality if $x \notin \Gamma$.*

Let $\{x_k\}_{k=1}^{\infty}$ be a generated by

$$x_{k+1} \in A(x_k).$$

If all points x_k are contained in a compact set $S \subset X$, then any limit of a convergent subsequence is a minimizer of f.

What kind of questions to expect?

In general, it is the instructor's job to design test questions in such a way that a well-prepared student can solve them in the available time. So if the instructor does his job correctly, you will not be asked to find the eigenvalues and eigenvectors of a 6 by 6 matrix or to minimize a complicated function of 10 variables.

A large portion of the test (more than half) will consist of routine questions designed to test your grasp of basic material. These could include questions like the ones below.

- Find the global minimum of the function $f = \ldots$ in the set $\Omega = \ldots$.
- Check whether the function $f = \ldots$ is convex in the set $\Omega = \ldots$.
- Suppose that $f = \ldots$ (a function of a single variable), and that you want to minimize f by Newton's method. Give an explicit formula for x_{n+1} in terms of x_n.
- State the Global Convergence Theorem
- Define a point to set mapping $A(x) = \ldots$.
  Is A closed?
- Suppose that $f = \ldots$ (a function of several variables), and that you want to minimize f by the method of steepest descent. If $x_0 = ...$, find x_1.
- Suppose that $f = \ldots$ (a function of a several variables), and that you want to minimize f by the method of steepest descent. Give an explicit formula for x_{n+1} in terms of x_n.
- Suppose that you want to minimize $f = \ldots$ (a quadratic function) by the method of steepest descent. If $x_0 = \ldots$, and if you know that the minimum value of f is $\ldots$, then how many iterations of the method of steepest descent are needed to compute the minimizer to within an accuracy of $1/100$?
- Suppose that you are given a 2×2 matrix $Q = \ldots$. Find a vector d_1 that is Q-orthogonal to the vector $d_0 = \ldots$.
- Suppose that Q is a $n \times n$ matrix, and that d_0 and g_1 are linearly independent vectors in E^n.
  Give a formula for a number s such that $g_1 + sd_0$ is Q-orthogonal to d_0.
  Is it always possible to find a number t such that $d_0 + tg_1$ is Q-orthogonal to d_0?

practice questions

One way to practice for the test is just to make up concrete instances of the kinds of basic questions described above.

For example:

Minimize a function

- Find the global minimum, if it exists, of the function
$$f(x, y) = x^2 + 4y^2 + 4z^2 - 2yz + 3x + 6y.$$
Justify your answer
- Find the global minimum, if it exists, of the function
$$f(x, y) = x^2 + 4y^2 - 4z^2 - 2yz + 3x + 6y.$$
Justify your answer.
- Find the global solution, if it exists, of the problem:
$$\text{minimize} \quad f(x, y) = x^3 + 2y^2 + 4xy + 4x + 4y \quad \text{in the set} \quad \{(x, y) \in E^2 : x \geq 0\}.$$
- find the global solution, if it exists, of the problem:
$$\text{minimize} \quad f(x, y) = x^3 + 2y^2 + 2xy + 4x + 4y \quad \text{in the set} \quad \{(x, y) \in E^2 : x \geq 0\}.$$
- Find the global solution, if it exists, of the problem:
$$\text{minimize} \quad f(x, y) = x^2 + 2y^2 + 2xy - 4x - 4y \quad \text{in the set} \quad \{(x, y) \in E^2 : x + y \geq 1\}.$$
- Find the global solution, if it exists, of the problem:
$$\text{minimize} \quad f(x, y) = x^2 + 2y^2 + 2xy - 4x - 4y \quad \text{in the set} \quad \{(x, y) \in E^2 : x + y \geq 1\}.$$

Clearly, you can make up more of these if you like.

Convexity:

Which of the following functions is convex?

- The function $f(x, y) = xy$ on the set $\Omega = \{(x, y) \in E^2 : x \geq 0, y \geq 0\}$.
- The function $f(x, y) = \frac{1}{xy}$ on the set $\Omega = \{(x, y) \in E^2 : x \geq 0, y \geq 0\}$.
- The function $f(x, y) = -\log(xy)$ on the set $\Omega = \{(x, y) \in E^2 : x \geq 0, y \geq 0\}$.
- The function $f(x, y) = -\log(x+y)$ on the set $\Omega = \{(x, y) \in E^2 : x + y \geq 0\}$.
- The function $f(x, y) = (1 - x^2)^2 + y^2$ on the set $\Omega = E^2$.

Of course, it is surely also possible, but harder, to think of a question about convexity that does not just ask you to check whether some function is convex.

Algorithms for minimizing a function of a single variable

- Let $f(x) = x^4 + e^x$ for $x > 0$. Suppose that you want to use Newton's method to compute a minimizer of f. Write down the formula for x_{k+1} in terms of x_k.
- more or less the same question with a different function f. (Make one up if you want practice.)

Global Convergence Theorem

- What is the definition of a *closed* point-to-set mapping?
- Define a point-to-set mapping A on E^n by

$$A(x) = \{y \in E^n : y^T x \leq 1\}.$$

  Show that A is closed at $x = 0$.
- Suppose that you want to minimize a C^1 function f no E^n. Consider the algorithm $x_{n+1} = x_n - \nabla f^T(x_n)$. Does this algorithm satisfy the hypotheses of the Global Convergence Theorem for every f?

  (This question is a bit hard for a test, because you have to first figure out whether you think the answer is yes or no, and it's not clear how to start thinking about it. But it's probably good practice.)

Steepest descent, condition number, etc

- Consider the 2×2 matrix

$$Q = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

  and assume that $a < c$.

  a. Let

$$d_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

  and compute $d_0^T Q d_0$.

  b. recall that the condition number of a matrix is defined to be

$$\text{condition number} \ = \frac{\text{largest eigenvalue}}{\text{smallest eigenvalue}}.$$

  Show that the condition number of Q is at least c/a.
- Suppose that f is a quadratic function on E^n:

$$f(x) = \frac{1}{2} x^T Q x + b^T x$$

  where Q is a symmetric $n \times n$ matrix and $b \in E^n$. Assume also that d is a nonzero vector in E^n, and define $g(s) = f(x + sd)$, for $s \in \mathbb{R}$.

  Derive the formula for the value s^* that minimizes g.
- Define a function f on E^2 by

$$f(x, y) = x^2 + xy + y^2 - 3x + y.$$

  If you try to find a minimizer for f using the method of steepest descent, and if your initial guess is $(x_0, y_0) = (5, 5)$, then write down the minimization problem that you would have to solve to find the point (x_1, y_1)

  or
- Define a function f on E^2 by

$$f(x, y) = x^2 + xy + y^2 - 3x + y.$$

  If you try to find a minimizer for f using the method of steepest descent, and if your initial guess is $(x_0, y_0) = (5, 5)$, What is the point (x_1, y_1)?

Note that last question is basically just a more concrete version of the earlier question ("derive the formula for the value s^* that minimizes g"). You may find that the abstract version is easier to solve. Of course, since you are not asked to derive anything in the last question, there you have the option of just writing down the correct answer with no proof, if you remember the formula. If you don't remember the formula, the easiest way to solve the last question may be to derive the general formula, and then to write down the answer – this way is cleaner, and there is less risk of making a computational mistake or getting bogged down in lots of numbers.

Conjugate directions, conjugate gradient

- Let
$$Q = \begin{pmatrix} 3 & 5 \\ 5 & 11 \end{pmatrix}.$$
Find a nonzero vector $d = (d_1, d_2)$ that is Q-orthogonal to $(-2, 3)$.
- Assume that Q is a symmetric, positive definite $n \times n$ matrix, and that d_0 and g_1 are linearly independent vectors in E^n.
Give a formula for a real number s with the property that
$$g_1 + sd_0 \quad \text{is } Q\text{-orthogonal to } d_0.$$

What *other* kinds of questions to expect?

There will also certainly be some questions (fewer than half) that are not as straightforward as most of the ones above, and should require a deeper understanding of the material.

The best way to prepare for these is to understand the material as well as possible.