Final Examination
Numerical Optimization MAD 5420-Spring 2004
Prof I.M.Navon
Time: 2 hours

Please solve 2 out of the 4 application questions as well as 2 out of 4 theoretical questions:

Applications

Numerical part

Solve 2 out of 4 questions

1. Consider the steepest descent method (p_k is the search direction) $p_k = -\nabla f(x_k)$ with exact line search to solve

$$min \quad f(x_1, x_2) = 4{x_1}^2 + 2{x_2}^2 + 4x_1x_2 - 3x_1$$

from the point $(2,2)^T$. Perform 3 iterations.

Show that the exact line search is giving the a stepsize α_k, such that,

$$\alpha_k = \frac{-\nabla f(x_k)^T p_k}{{p_k}^T Q p_k}.$$

2. Consider the Quasi-Newton method BFGS. Let

$$s_k = x_{k+1} - x_k$$

and

$$y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$$
$$B_{k+1} = B_k - \frac{(B_k s_k)(B_k s_k)^T}{{s_k}^T B_k s_k} + \frac{y_k {y_k}^T}{{y_k}^T s_k}$$

minimize

$$f(x_1, x_2) = x_1 - x_2 + 2{x_1}^2 + 2x_1x_2 + {x_2}^2$$

from the starting point $\underline{x_0} = (0,0)^T$.

Use

$$B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Carry out 3 iterations.

Hint:

To find the minimizing step length α^* along s_1, minimize $f(x_1 + \alpha_k s_1)$, STOP if $\|\nabla f_k\|_2 < \epsilon$. Take ϵ to be equal to 0.01.

3. Consider the augmented Lagrangian method for equality constrained minimization.

$$min\ \mathcal{L}(x, \lambda, \rho) = f(x) - \lambda^T g(x) + \frac{1}{2}\,\rho\ g(x)^T g(x)$$

to minimize $f(X)$ such that $g(x) = 0$. Here we use the necessary conditions $\frac{\partial f}{\partial x_i}$, $i = 1, 2$ to solve the unconstrained subproblem.

Update using

$$\lambda_{k+1} = \lambda_k + \rho_k\ g(x_k)$$

$$\rho_{k+1} = \beta\ \rho_k, \beta > 1$$

Start with $\lambda_0 = 0$, $\rho_k = 0.1\ 8^k$ to minimize

$$f(x_1, x_2) = \frac{1}{2}({x_1}^2 + \frac{1}{3}{x_2}^2)$$

such that

$$g(x) = x_1 + x_2 - 1 = 0$$

Carry out 3 minimizations. Show that the use of first order necessary conditions leads to

$${x_1}^{(k)} = \frac{\rho_k - \lambda_k}{1 + 4\ \rho_k}$$

$${x_2}^{(k)} = \frac{3\ (\rho_k - \lambda_k)}{1 + 4\ \rho_k}.$$

4. Consider the penalty method of quadratic penalty type.

$$min\ f(x_1, x_2) = -x_1\ x_2$$

such that $g(x) = x_1 + 2\,x_2 - 4 = 0$.

i.e.,

$$min\ \mathcal{P}(x, \rho) = -x_1\,x_2 + \frac{1}{2}\,\rho(x_1 + 2\,x_2 - 4)^2$$

ρ is the penalty parameter. Use necessary conditions of unconstrained minimization problem to show that for $\rho > \frac{1}{4}$,

$$x_1(\rho) = x_1 = \frac{8\ \rho}{4\ \rho - 1}$$

$$x_2(\rho) = x_2 = \frac{4\ \rho}{4\ \rho - 1}$$

and $x^{opt} = (2, 1)^T$.

Compute the condition number of the Hessian $\nabla_x^2 \mathcal{P}(x, \rho)$ at $X(\rho)$ and show that it is approximately equal to $\frac{25\,\rho}{4}$.

On the basis of this result, comment on the ill-conditioning of the penalty method when $\rho \to \infty$.

Theoretical part

Solve 2 out of the following 4 questions concerning theoretical aspects of numerical optimization:

1.

a. Describe the algorithmic form of the Davidon-Fletcher-Powell Quasi-Newton algorithm.

b. Prove the hereditary positive definiteness of DFP, i.e., if H_k is positive definite so is H_{k+1} i.e. show that:

$x^T H_{k+1} x > 0$ for all $x \neq 0$

Hint: The DFP quasi Newton update formula is given by:

$$H_{k+1} = H_k + \frac{p_k p_k^T}{p_k^T y_k} - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k}$$

$$p_k = x_{k+1} - x_k$$
$$y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$$

2. Prove the conjugate-gradient theorem, i.e., that for the conjugate- gradient algorithm for the quadratic problem: min $\frac{1}{2}x^TQx - b^Tx$:

$$d_0 = -g_0 \text{ or } b - Qx_0$$

$$\alpha_k = \frac{-g_k^T d_k}{d_k^T Q d_k}$$

$$d_{k+1} = -g_{k+1} + \beta_k d_k$$

$$\beta_k = \frac{g_{k+1} Q d_k}{d_k^T Q d_k}$$

is a conjugate direction method. If it does not terminate at x_k then:

$$a)[g_0, g_1, \ldots, g_k] = [g_0, Qg_0, \ldots, Q^k g_0]$$

$$b)[d_0, d_1, \ldots, d_k] = [g_0, Qg_0, \ldots, Q^k g_0]$$

$$c)d_k Q d_i = 0 \text{ for } i \leq k-1$$

$$d)\alpha_k = \frac{g_k^T g_k}{d_k^T Q d_k}$$

$$e)\beta_k = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k}$$

3. Prove the second order necessary conditions Theorem for equality constraints. The Theorem states: Suppose x^* is a local minimum of the function f subject to the constraints

$$h(x) = 0$$

and x^* is a regular point of these constraints. Then there exists a $\lambda \epsilon E^m$ such that:

$$\nabla f(x^*) + \lambda^T \nabla h(x^*) = 0.$$

If we denote by M the tangent plane:

$$M = \{y : \nabla h(x^*)y = 0\}$$

Then the matrix:

$$L(x^*) = G(x^*) + \lambda^T H(x^*)$$

where $H = \nabla^2 h_i(x^*)$ and G is the Hessian of the function f, is positive semi-definite on M for all $y \epsilon M$. That is:

$$y^T L(x^*) y \geq 0$$

for all $y \in M$.

Here

$$L = G + \lambda^T H$$

is the matrix of second order partial derivatives, with respect to x of the Lagrangian

$$l(x, \lambda) = f(x) + \lambda^T h(x)$$

4. Describe the algorithm for sequential quadratic programming (SQP).

It is used as a generalization of Newton's method for unconstrained minimization by obtaining a search direction by solving a problem with quadratic objective function and linear constraints

Derive it from the basic problem:

Minimize f(x)

Subject to g(x)=0

Write Lagrangian for problem and first optimality conditions and write formula for Newton' s method updating direction and Lagrange multipliers.

Show how you obtain the quadratic program:

$$\begin{pmatrix} \nabla^2_{xx} L & -\nabla g \\ -\nabla g^T & 0 \end{pmatrix} \begin{pmatrix} p \\ \nu \end{pmatrix} = \begin{pmatrix} -\nabla_x L \\ g \end{pmatrix}$$

And the new estimates of the solution. Describe the SQP algorithm.

b. Since the progress of the SQP method is measured using merit functions prove the following theorem regarding descent direction for the merit function:

Assume that (p_k, ν_k) is computed as solution of the quadratic program

Minimize

$$\text{p} \qquad \frac{1}{2} p^T H p + p^T [\nabla_x L(x_k, \lambda_k)]$$

subject to $[\nabla g(x_k)]^T p + g(x_k) = 0.$

where H is some positive definite approximation to $\nabla^2_{xx} L(x_k, \lambda_k)$. If $p_k \neq 0$ then

$p_k{}^T \nabla \mathrm{M}(x_k) < 0$

For all sufficiently large values of penalty parameter ρ where

M(x) = $f(x) + \rho g(x)^T g(x)$

That is, p_k is a descent direction with respect to this merit function.