Numerical Optimization Course H03E3a Exam - REPRINT

Moritz Diehl
Optimization in Engineering Center OPTEC and Department ESAT, KU Leuven, Belgium

January 16, 2013, 8:00-10:00, C200

page	0	1	2	3	4	5	6	7	8	9	sum
points on page (max)	5	10	7	11	12	4	11	12	14	4	90
points obtained											
intermediate sum											
intermediate sum (max)	5	15	22	33	45	49	60	72	86	90	

Total Mark	Project Mark	Project Defense	Exam Defense	Exam Mark

NAME OF STUDENT:..

On the following sheets you find 15 questions on 10 pages with altogether 90 points. You may use the space below the questions for the answers, or use extra sheets or the back-sides, always clearly stating the question number and subitem letter (e.g. "3.(a)"). To keep in time, you may at first skip those questions that you find difficult. If you take one minute per point (e.g. 2 minutes for question 1) you will need 1.5 hours for writing the exam, safely in time. At the end, please return both the exam sheet and all your extra papers stapled together. You will later discuss your results with the examiner for 10 minutes.

Good luck!

1. Define mathematically what is a global minimizer x^* of the problem

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad x \in \Omega$$

2	

2. Compute gradient $\nabla f(x)$ and Hessian $\nabla^2 f(x)$ of the function $f : \mathbb{R}^2 \to \mathbb{R},\ (x_1, x_2) \mapsto f(x_1, x_2) := x_1 x_2$.

3	

total on page: 5	

3. (Duality) Regard the following general NLP (the *primal problem*):

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad \begin{cases} g(x) & = & 0 \\ h(x) & \geq & 0 \end{cases}$$

(a) Define the Lagrangian function $\mathcal{L}(x, \lambda, \mu)$ of the general NLP.

2	

(b) Define the Lagrange dual function $q(\lambda, \mu)$

2	

(c) Is the Lagrange dual function $q(\lambda, \mu)$ convex or concave or nothing of the two? Justify your answer.

2	

(d) State the dual problem.

2	

(e) What is weak duality? Does it hold for a general (possibly non-convex) primal NLP?

2	

total on page: 10	

4. Line search: regard an iterative descent algorithm for unconstrained minimization of a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, at the current iterate x_k and with a search direction p_k.

   (a) When does a step length t satisfy the so called "Armijo-Condition"?

3	

   (b) Describe the backtracking algorithm to find a point that satisfies the Armijo-condition.

2	

5. Convergence rates: You observe an iterative optimization algorithm while it converges towards a solution. In each iteration, it gives you the norm of the current gradient. You see the sequence

```
iter     |gradient|
k=0        3.16302341E-1
k=1        1.00000011E-1
k=2        1.00000027E-2
k=3        1.00000027E-4
.
.
.
k=?        1.00000000E-16
**** convergence achieved ****
```

   What local convergence rate seems this algorithm to have? At what iteration counter `k` would the desired accuracy of 10^{-16} be reached?

2	

total on page: 7	

6. Define what it means that a set $\Omega \subset \mathbb{R}^n$ is convex?

3	

7. Constrained Optimization: What are the local convergence rates of

- the SQP Method with BFGS Hessian Updates ?
- the Constrained Gauss-Newton method ?

2	

8. Regard the numerical solution of the constrained nonlinear least-squares problem

$$\min_{x \in \mathbb{R}^n} \frac{1}{2}\|F(x)\|_2^2 \quad \text{subject to} \quad g(x) \;=\; 0$$

with $F : \mathbb{R}^n \to \mathbb{R}^{n_F}$ and $g : \mathbb{R}^n \to \mathbb{R}^{n_g}$.

(a) Write down the Lagrangian function $\mathcal{L}(x, \lambda)$ of this problem.

1	

(b) Write down the Hessian $\nabla_x^2 \mathcal{L}(x, \lambda)$ of the Lagrangian, using the expressions $J(x) = \frac{\partial F}{\partial x}(x)$, $\nabla_x^2 F_i(x)$ for $i = 1, \ldots, n_F$ and $\nabla_x^2 g_j(x)$ for $j = 1, \ldots, n_g$.

3	

(c) Underline which of the above terms are neglected in the Gauss-Newton Hessian approximation B_{GN} (which is similar but not equal to $\nabla_x^2 \mathcal{L}(x, \lambda)$)

1	

(d) When does the constrained Gauss-Newton method converge fast?

1	

total on page: 11	

9. Which of the following sets Ω is convex ? Encircle the convex sets.

   (a) $\Omega = \{x \in \mathbb{R}^n \mid x^T x \leq 10\}$

   (b) $\Omega = \{x \in \mathbb{R}^n \mid \|x\|_2 \leq 10\}$

   (c) $\Omega = \{x \in \mathbb{R}^n \mid \|x\|_2 \geq 10\}$

   (d) $\Omega = \{x \in \mathbb{R}^n \mid a^T x \geq 10\}$

   (e) $\Omega = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n \mid \|y\|_2^2 \leq 10 + x^T y - \|x\|_2^2\}$

5	

10. Regard the non-smooth optimization problem

$$\min_{x \in \mathbb{R}^n} \quad \|F(x)\|_1$$

with a differentiable function $F : \mathbb{R}^n \to \mathbb{R}^m$. Formulate this problem into an equivalent smooth nonlinear program, using an appropriate number of slack variables s_i if necessary.

3	

11. Automatic differentiation in forward and backward mode: regard the task to compute the gradient of a scalar function $f : \mathbb{R}^n \to \mathbb{R}$. If evaluating $f(x)$ uses one second of CPU time and $n = 30$, how much time do you need to compute $\nabla f(x)$ using the forward and how much using the backward mode mode of automatic differentiation? What is the disadvantage of the backward mode compared to the forward mode ?

4	

total on page: 12	

12. Automatic Differentiation in Backward Mode: regard the following algorithm to evaluate the function $f : \mathbb{R}^{30} \to \mathbb{R}$.

```
function [f]=myfunction(u)
x(1)=1;
for i=1:30;
 x(i+1)=x(i)*u(i);
end
f=x(31)*x(30);
```

We now write an algorithm that computes the gradient $\nabla f(x)$ by the backward mode of automatic differentiation. Please add the missing three lines to the following template function. Remember that the meaning of `Xbar` is $\frac{df}{dX}$ for any variable X used in the code.

```
function [f,ubar]=mygradient(u)
% start with forward evaluation
 x(1)=1;
 for i=1:30
  x(i+1)=x(i)*u(i);
 end
 f=x(31)*x(30);
% initialize nearly all adjoint variables by zero
 xbar=zeros(31,1);

 ubar= ...

 fbar =1;

% backwards sweep, in reverse order for each line
%  (f=x(31)*x(30);)
 xbar(31)=xbar(31)+fbar*x(30);
 xbar(30)=xbar(30)+fbar*x(31);

% reverse the loop
for i=30:-1:1
% (x(i+1)=x(i)*u(i);)

...

...

end
```

4	

total on page: 4	

13. A three page question: Regard the following optimization problem (note that it is a maximization problem):

$$\max_{x \in \mathbb{R}^2} x_2^2 - x_2 \quad \text{subject to} \quad \begin{cases} x_1^2 + x_2^2 & \leq & 4 \\ x_1 & \geq & -1 \\ x_1 & \leq & 1 \end{cases}$$

(a) How many variables, how many equality, and how many inequality constraints does this problem have?

3	

(b) Sketch the feasible set Ω of this problem.

3	

(c) Bring this problem into the NLP standard form (a minimization problem):

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad \begin{cases} g(x) & = & 0 \\ h(x) & \geq & 0 \end{cases}$$

by defining the dimension n and the functions f, g, h along with their dimensions appropriately

3	

FROM NOW ON UNTIL THE END TREAT THE PROBLEM IN THIS STANDARD FORM.

(d) An optimal solution of the problem is $x^* = (0, 2)^T$. What is the active set $\mathcal{A}$ at this point?

2	

total on page: 11	

(e) Is the *linear independence constraint qualification (LICQ)* satisfied at x^* ?

3	

(f) Write down the Lagrangian function of this optimization problem

2	

(g) Formulate the necessary optimality conditions of first order (also called Karush-Kuhn-Tucker (KKT) conditions) that a local minimizer $x^* \in \mathbb{R}^2$ of this problem must satisfy, both generically and specifically.

4	

(h) Describe the tangent cone $T_\Omega(x^*)$ (the set of feasible directions) to the feasible set at this point x^*, both by a set definition formula **and by a sketch**

3	

total on page: 12	

(i) Find multiplier vectors λ^*, μ^* so that the above point x^* satisfies the KKT conditions

3	

(j) Describe the critical cone $C(x^*, \mu^*)$ at the point (x^*, μ^*) both in a formula **and a sketch**

3	

(k) Formulate the second order sufficient conditions for optimality (SOSC) for this problem

3	

(l) Prove that the point x^* is a local minimizer

3	

(m) Is the point x^* also a global minimizer? Justify

2	

total on page: 14	

14. Discrete Time Optimal Control: Regard the dynamic system $x_{k+1} = f(x_k, u_k)$ with $k = 0, \dots, N-1$. How many degrees of freedom does the NLP generated in the **sequential** (or **direct single shooting**) method with free initial value have, if $x \in \mathbb{R}^{n_x}$ and $u \in \mathbb{R}^{n_u}$, and if you introduce N control intervals? Encircle the correct answer

   (a) $(n_x + n_u)N + n_x$
   (b) $(1 + n_x)N$
   (c) $n_u N + n_x$
   (d) $n_u + N$

2	

15. Regard again the dynamic system $x_{k+1} = f(x_k, u_k)$ with $k = 0, \dots, N-1$. How many degrees of freedom does the NLP generated in the **simultaneous** (or **direct multiple shooting**) method have, if $x \in \mathbb{R}^{n_x}$ and $u \in \mathbb{R}^{n_u}$, and if you introduce N control intervals? Encircle the correct answer

   (a) $(n_x + n_u)N + n_x$
   (b) $(1 + n_x)N$
   (c) $n_u N + n_x$
   (d) $n_u + N$

2	

total on page: 4	