Final Preparation Handout

CS/SWE 4/6TE3 December 2, 2010

Review of Previous Final Exams

Solution of Final Exam 2004

1. Concepts, convexity, duality:

a. What is the rate of convergence of the following sequence? Justify your answer.

$$x^i = \frac{1}{\log(i+1)} \qquad i = 1, 2, \ldots$$

b. Show that the function $f(x) : R^2 \to R$ given by

$$f(x) := e^{x_1^4 + x_2^2 - 2x_1 - x_2 - 1}$$

is a convex function.

c. Consider the nonlinear optimization problem:

$$\begin{array}{llll} \text{(NLO)} & \min & f(x) & \\ & \text{s.t.} & g_j(x) & \le 0, \quad j = 1, \ldots, k, \\ & & Ax & = b, \\ & & x & \in R^n, \end{array}$$

where the functions $R^n \to R$ are convex, $A : m \times n$ is a matrix and $b \in R^m$.
Define when a point $x^0 \in R^n$ is a Slater point of this problem.

d. Describe how one can verify if problem (NLO) satisfies the Slater condition. Justify your procedure.

e. Give the Lagrange Function and the Wolfe Dual for this problem.

f. Derive the KKT conditions of the given (NLO) problem.

12p

Solution

a. Since

$$\lim_{i\to\infty} x^i = 0,$$

$$\lim_{i\to\infty} \frac{x^{i+1} - 0}{x^i - 0} = \lim_{i\to\infty} \frac{\log(i+1)}{\log(i+2)} = \lim_{i\to\infty} \frac{\frac{1}{i+1}}{\frac{1}{i+2}} = 1,$$

and

$$\lim_{i\to\infty} \frac{x^{i+1} - 0}{(x^i - 0)^{1+\epsilon}} = \lim_{i\to\infty} \frac{x^{i+1}}{x^i} \cdot \log^\epsilon(i+1) = \infty,$$

the rate of convergence is 1, and is sublinear.

b. $x_1^4, x_2^2, -2x_1 - x_2 - 1$ are all convex, so the sum of them is convex; e^x is a convex and increasing function; as a composition of these two functions, the given function is therefore convex.

c. Denote I_L/I_N the sets of indices of all linear/nonlinear inequalities. $x^0 \in \mathbb{R}^n$ is a Slater point of this problem iff it satisfies all of the following conditions

$$g_j(x^0) \le 0, \forall j \in I_L \tag{1a}$$

$$g_j(x^0) < 0, \forall j \in I_N \tag{1b}$$

$$Ax^0 = b \tag{1c}$$

d. We can solve the following problem to verify if problem (NLO) satisfies the Slater condition:

$$\begin{aligned} \min \quad & \tau \\ \text{s.t.} \quad & g_j(x) - \tau \leq 0, \quad \forall j \in I_N \\ & g_j(x) \leq 0, \qquad \forall j \in I_L \\ & Ax = b, \\ & x \in R^n, \tau \in R \end{aligned}$$

If the optimum $\tau^* > 0$, the (NLO) problem is infeasible;
if $\tau^* = 0$, there is no Slater point for the (NLO) problem;
if $\tau^* < 0$, the (NLO) problem satisfies Slater condition.

e. The Lagrange function of this problem is

$$L(x, \lambda, \mu) = f(x) + \sum_{j=1}^{k} \lambda_j g_j(x) + \mu^T (b - Ax), \quad \lambda_j \geq 0, j = 1, \cdots, k.$$

The Wolfe dual of this problem is

$$\begin{aligned} \sup_{x,\lambda,\mu} \quad & L(x, \lambda, \mu) \\ \text{s.t.} \quad & \nabla_x L(x, \lambda, \mu) = 0 \\ & \lambda_j \geq 0, j = 1, \cdots, k. \end{aligned}$$

Substituting $L(x, \lambda, \mu)$ we get

$$\begin{aligned} \sup_{x,\lambda,\mu} \quad & f(x) + \sum_{j=1}^{k} \lambda_j g_j(x) + \mu^T (b - Ax) \\ \text{s.t.} \quad & \nabla f(x) + \sum_{j=1}^{k} \lambda_j \nabla g_j(x) - A^T \mu = 0 \\ & \lambda_j \geq 0, j = 1, \cdots, k. \end{aligned}$$

f. The KKT condition for this problem is

$$\begin{aligned} g_j(x) &\leq 0, \quad j = 1, \ldots, k, \\ Ax &= b, \\ \lambda_j &\geq 0, \quad j = 1, \cdots, k, \\ \nabla_x L(x, \lambda, \mu) &= 0, \\ \lambda_j g_j(x) &= 0, \quad j = 1, \cdots, k. \end{aligned}$$

Substituting $L(x, \lambda, \mu)$ we get

$$\begin{aligned} g_j(x) &\leq 0, \quad j = 1, \ldots, k, \\ Ax &= b, \\ \lambda_j &\geq 0, \quad j = 1, \cdots, k, \\ \nabla f(x) + \sum_{j=1}^{k} \lambda_j \nabla g_j(x) - A^T \mu &= 0, \\ \lambda_j g_j(x) &= 0, \quad j = 1, \cdots, k. \end{aligned}$$

---

2. Unconstrained Optimization: Consider the function

$$f(x) = x_1 + x_2 - \frac{1}{|x_1 + x_2|^2 + 1}.$$

a. Is $f(x)$ a convex function?

b. Let $\bar{x}^T = (-1, 1)$. Which algorithm would be the best to minimize the function $f(x)$ starting from the given point $\bar{x}$? Give at least two arguments why would you choose that algorithm.

c. Apply two steps of the Polak-Ribière Conjugate Gradient algorithm starting from $x^0 = (-1, 1)^T$. (Use an approximate line-search.)

d. Let $x^1 = (0.5, -0.5)^T$, $x^2 = (0.5, 0.5)^T$ and $x^3 = (-0.5, 0.5)^T$. Apply one step of the Nelder-Mead Simplex Method with $\alpha = 1/2,\ \ \beta = 1/3,\ \ \gamma = 2$.

e. Give the interpolation polynomial $Q(x)$ of the function $f(x)$ through the points $x^1,\ x^2,\ x^3$ by using the basis functions $\phi_1(x) = 1,\ \phi_2(x) = x_1x_2$ and $\phi_3(x) = 2x_1 + x_2$.

10p

Solution

a. $f(x)$ is not convex. For $x^1 = (2, 2), x^2 = (4, 4)$,

$$\begin{aligned} f(\frac{x^1 + x^2}{2}) &= f(3, 3) = 5.9730 \\ &> \frac{f(x^1) + f(x^2)}{2} = 5.9629, \end{aligned}$$

which contradicts the definition of convex functions.

b. Any algorithm other than Newton can be used. For example, Trust Region (1) could have fast convergence rate, nearly Newton, and (2) allows to avoid line search.

c. $g = \nabla f(x) = (1 + 2\frac{x_1+x_2}{(x_1+x_2)^2+1)^2}, 1 + 2\frac{x_1+x_2}{((x_1+x_2)^2+1)^2})^T$,
$x^0 = (-1, 1)^T, f^0 = -1, g^0 = (1, 1)^T$.
$s^0 = -g^0 = (-1, -1)^T$. $f(x^0 + s^0) = f(-2, 0) = -2.2000 < f^0$; choose step length 1.
$x^1 = (-2, 0), f^1 = -2.2000, g^1 = (0.8400, 0.8400)^T$.
$s^1 = -g^1 + \frac{g^{1T}(g^1-g^0)}{g^{0T}g^0}s^0 = (-0.7056, -0.7056)^T$. $f(x^1 + s^1) = f(-2.7056, -0.7056) = -3.4903 < f(x^1)$; choose step length 1.
$x^2 = (-2.7056, -0.7056), f^2 = -3.4903, g^2 = (0.9573, 0.9573)^T$.

d. Sort the points in the order of ascending function value

$$f(x^0) = f(0.5, -0.5) = -1 \leq f(x^1) = f(-0.5, 0.5) = -1 \leq f(x^2) = f(0.5, 0.5) = 0.5;$$

$\bar{x} = \frac{x^0+x^1}{2} = (0, 0)^T$; $x^r = \bar{x} + (\bar{x} - x^2) = (-0.5, -0.5)^T$;
$f(x^r) = -1.5 < f(x^0)$;
$x^e = \bar{x} + \gamma(\bar{x} - x^2) = (-1, -1)^T$; $f(x^e) = -2.2 < f(x^r)$;
Drop $x^2 = (0.5, 0.5)^T$ and add $x^e = (-1, -1)^T$.

e. Suppose the coefficient of ϕ_i is a_i. Then $a = (a_1, a_2, a_3)^T$ is given by the solution of the following linear system

$$\begin{pmatrix} 1 & -0.25 & 0.5 \\ 1 & 0.25 & 1.5 \\ 1 & -0.25 & -0.5 \end{pmatrix} a = \begin{pmatrix} -1 \\ 0.5 \\ -1 \end{pmatrix}$$

The solution is $(-0.25, 3, 0)$. The polynomial required is $-0.25 + 3x_1x_2$.

---

3. Linearly Constrained NLO – Modelling: *Tartaglia's (1500-1557) problem:* How to divide the number 8 into two parts such that the result of multiplying the product of the two parts by their difference will be maximal.

a. Model Tartaglia's problem as a constrained NLO problem.

b. Prove that both parts are nonzero at the optimal solution.

c. Either show that your model is convex or give an equivalent convex optimization problem.

d. Prove that the point $x = (4+4/\sqrt{3}, 4-4/\sqrt{3})$ is the optimal solution for Tartaglia's problem.

8p

Solution Denote the two parts as x and y.

a. The problem can be formulated as follows

$$\begin{aligned} \max_{x,y} \quad & xy|y-x| \\ \text{s.t.} \quad & x+y=8 \end{aligned}$$

b. For both $(8,0)$ and (0,8), the objective function value are zero, while for $(3,5)$ the objective function value is 30. So neither x nor y is zero at optimal solution.

c. If either x or y is negative, the other has to be positive, thus the objective function value is negative. Therefore, we know that both x and y are nonnegative at optimal solution. Considering the symmetry between x and y, we could assume $y \geq x$ without loss of generality. The problem then is equivalent to

$$\begin{aligned} \max_{x,y} \quad & xy(y-x) && \text{(2a)} \\ \text{s.t.} \quad & x+y=8 && \text{(2b)} \\ & y \geq x \geq 0 && \text{(2c)} \end{aligned}$$

Since $\log t$ is an increasing function, problem (2) is equivalent to

$$\begin{aligned} \max_{x,y} \quad & \log x + \log y + \log(y-x) && \text{(3a)} \\ \text{s.t.} \quad & x+y=8 && \text{(3b)} \\ & y \geq x \geq 0. && \text{(3c)} \end{aligned}$$

All the three addends of the objective function are concave, and so the objective function itself. Problem (3) is then a convex formulation of Tartaglia's problem.

d. Substitute $y = 8 - x$ into (2a), we get the objective function

$$f(x) = 2x(8-x)(8-2x). \tag{4}$$

The constraints becomes

$$0 \leq x \leq 4.$$

Since $f(0) = f(4) = 0$, the maximum value cannot be attained at the boundary. Hence the maximum value has to be attained at a stationary point of $f(x)$. Considering

$$f'(x) \equiv 2(2x^2 - 24x + 32) = 0 \Leftrightarrow x = \frac{24 \pm \sqrt{24^2 - 4 \cdot 3 \cdot 32}}{6} = 4 \pm \frac{4}{\sqrt{3}}$$

and $0 \leq x \leq 4$, the optimal x is $4 - \frac{4}{\sqrt{3}}$, and the corresponding y is $8 - x = 4 + \frac{4}{\sqrt{3}}$.

---

4. Linearly Constrained NLO : Consider the following linearly constrained non-linear optimization (NLO) problem:

$$\text{(NLO)} \qquad \begin{array}{ll} \min & x_1 + x_2 + x_1 x_2 \\ \text{s.t.} & 3x_1 + 2x_2 = 5. \end{array}$$

a. Reduce this problem to an unconstrained optimization problem.

b. Make a Quasi-Newton step with the DFP update for the *reduced* problem from **a**. The current point is $x^1 = (1,1)^T$ and the current approximate of inverse Hessian is two times the identity matrix.

c. Let us assume that the variables x_1 and x_2 are nonnegative. Make one step of the Reduced Gradient algorithm from the point $(1,1)$ when x_2 is the basis variable.

8p

Solution

a. Substituting $x_2 = (5-3x_1)/2$ into the objective function, we get the reduced unconstrained problem

$$\min \quad f_N(x_1) = \frac{5}{2} + 2x_1 - \frac{3}{2}x_1^2.$$

b. The gradient of $f_N(x_1)$ is $\nabla_N f(x_1) = 2 - 3x_1$. $x^1 = 1, g^1 = -1, H^1 = 2$; $s^1 = -H^1 g^1 = 2$; $f_N(1+2\lambda) = -6\lambda^2 - 2\lambda + 3$; no λ can satisfy Wolfe condition for any $0 < c_1 < c_2 < 1$, and the objective function approaches $-\infty$ when λ approaches ∞; so the minimization process is over, no DFP update needed.

c. $x^1 = (1,1)^T, g^1 = (2,2)^T, g_N^1 = (1, -3/2)g^1 = -1$.
$s_N^1 = 1$; $s_B^1 = -3/2 s_N^1 = -3/2$; $s^1 = (1, -3/2)$.
$\lambda_{\max} = 2/3$; $f(1+\lambda, 1-3\lambda/2) = -\frac{3}{2}\lambda^2 - \lambda + 3$; the minimum is attained at $\lambda^* = \lambda_{\max} = 2/3$; $x^2 = x^1 + \lambda^* s^1 = (1.6667, 0)^T$;

---

5. Non-linearly Constrained NLO: Consider the following constrained non-linear optimization (NLO) problem:

$$\begin{array}{llll} \text{(NLO)} & \min & e^{x_2} & \\ & \text{s.t.} & \sqrt{x_1^2+x_2^2} & \leq x_1+1. \end{array}$$

a. Is this a convex optimization problem? Justify your answer.

b. Give the Wolfe Dual of the (NLO) problem.

c. Give the SQP quadratic subproblem for the (NLO) problem at the point $x^1 = (1,1)^T$ and Lagrange multiplier value $y^1 = 2$. (Hint: to transform the inequality constraint into equality constraint you can add squared slack variable to the constraint.)

d. Using the logarithmic barrier function reformulate the constrained NLO problem as a series of parameterized unconstrained optimization problems.

12p

Solution

a. (NLO) is a convex optimization problem: $\sqrt{x_1^2+x_2^2} = \|(x_1,x_2)\|_2$ is convex, hence the constraint is convex; the objective function e^{x_2} is convex.

b. The Lagrange function of this problem is

$$L(x,\lambda) = e^{x_2} + \lambda(\sqrt{x_1^2+x_2^2} - x_1 - 1), \quad \lambda \geq 0$$

and

$$\nabla_x L(x,\lambda) = \left(\lambda\left(\frac{x_1}{\sqrt{x_1^2+x_2^2}} - 1\right), e^{x_2} + \frac{\lambda x_2}{\sqrt{x_1^2+x_2^2}}\right)^T,$$

so its Wolfe dual is

$$\begin{array}{ll} \sup\limits_{x,\lambda} & e^{x_2} + \lambda(\sqrt{x_1^2+x_2^2} - x_1 - 1) \\ \text{s.t.} & \lambda\left(\dfrac{x_1}{\sqrt{x_1^2+x_2^2}} - 1\right) = 0 \\ & e^{x_2} + \dfrac{\lambda x_2}{\sqrt{x_1^2+x_2^2}} = 0 \\ & \lambda \geq 0 \end{array}$$

c. We first add a squared slack variable $x_3 \in \mathbb{R}$ to the constraint as follows

$$\sqrt{x_1^2+x_2^2} + x_3^2 - x_1 - 1 = 0.$$

Lets choose $x_3 = 0.7654$ which is feasible for $(1,1)^T$, in this case we get (1,1,0.7654) as feasible starting point of the transformed problem. An infeasible point could also be used as starting point.

The Lagrange function of this transformed problem is

$$L(x,\lambda) = e^{x_2} + \lambda(\sqrt{x_1^2+x_2^2} + x_3^2 - x_1 - 1),$$

$$\nabla_x L(x,\lambda) = \left(\lambda\left(\frac{x_1}{\sqrt{x_1^2+x_2^2}} - 1\right), e^{x_2} + \frac{\lambda x_2}{\sqrt{x_1^2+x_2^2}}, 2\lambda x_3\right)^T,$$

$$\nabla_{xx} L(x,\lambda) = \begin{pmatrix} \frac{\lambda x_2^2}{(x_1^2+x_2^2)^{3/2}} & -\frac{\lambda x_1 x_2}{(x_1^2+x_2^2)^{3/2}} & 0 \\ -\frac{\lambda x_1 x_2}{(x_1^2+x_2^2)^{3/2}} & e^{x_2} + \frac{\lambda x_1^2}{(x_1^2+x_2^2)^{3/2}} & 0 \\ 0 & 0 & 2\lambda \end{pmatrix}.$$

For point $(1,1,0.7654)^T$ and $\lambda = 2$,

$$\nabla_{xx} L((1,1,0.7654),2) = \begin{pmatrix} 0.7071 & -0.7071 & 0 \\ -0.7071 & 3.4254 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

$$\nabla_x L((1,1,0.7654),2) = \begin{pmatrix} -0.5858 \\ 4.1325 \\ 3.0615 \end{pmatrix}.$$

The Jacobian of constraints at $(1,1,0.7654)^T$ is $\nabla H(x) = (-0.2929, 0.7071, 1.5307)$, and the function value of the constraint is 0. The SQP requested is then

$$\begin{aligned} \min_{\Delta x} \quad & \frac{1}{2}\Delta x^T \begin{pmatrix} 0.7071 & -0.7071 & 0 \\ -0.7071 & 3.4254 & 0 \\ 0 & 0 & 4 \end{pmatrix} \Delta x + \begin{pmatrix} -0.5858 \\ 4.1325 \\ 3.0615 \end{pmatrix}^T \Delta x \\ \text{s.t.} \quad & (-0.2929, 0.7071, 1.5307)\Delta x = 0 \end{aligned}$$

d. Using log-barrier function, this constrained optimization problem can be formulated as a series of the following unconstrained optimization problem

$$\min_x \quad e^{x_2} - \mu \log\left(x_1 + 1 - \sqrt{x_1^2 + x_2^2}\right)$$

where $\mu > 0$ is a parameter.

THE END