Computational Methods of Optimization
First Midterm(23rd Sep, 2015)

Answer all questions

Answer all questions Time: 90 minutes

1. Let $f(x_1, x_2) = x_1^4 + x_2^4 - 4x_1x_2$. where $x_1, x_2 \in \mathbb{R}$.
   (a) Is f coercive? Give reasons — 5 marks
   (b) Find global minimum of f? — 5 marks

2. Let $C \subseteq \mathbb{R}^n$ be a convex set. Consider the following problem

$$\min_{\mathbf{x} \in C} f(\mathbf{x})$$

   where f is not neccessarily $\mathcal{C}^1$. Let $\mathbf{x}_1 \in C$ and $\mathbf{x}_2 \in C$ arbe two distinct local minima of a convex function f. It is also given that $\mathbf{x}_1$ and $\mathbf{x}_2$ are interior points. Prove or Disprove: Global minimum of the function f is attained at $\mathbf{x} = \frac{1}{2}(\mathbf{x}_1 + \mathbf{x}_2)$. — 10 marks

3. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a $\mathcal{C}^1$ function. Let $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k \mathbf{u}^{(k)}$ be a sequence of points where $\mathbf{u}^{(k)}$ is a descent direction and α_k is the step-size. Show that for steepest descent direction with exact line search, $\mathbf{u}^{(k+1)\top}\mathbf{u}^{(k)} = 0$ holds — 5 marks

4. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a $\mathcal{C}^1$ function with the following property.

$$\text{For all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n \; f(\mathbf{y}) - f(\mathbf{x}) - \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) \leq \frac{L}{2}\|\mathbf{x} - \mathbf{y}\|^2$$

   where $L > 0$. Consider the following iterative algorithm for minimizing $f(\mathbf{x})$. Starting from $\mathbf{x}^{(0)} \in \mathbb{R}^n$, for $k \geq 0$

$$\mathbf{x}^{(k+1)} = min_{\mathbf{y}} g(\mathbf{y}) \; g(\mathbf{y}) = f(\mathbf{x}^{(k)}) + \nabla f(\mathbf{x}^{(k)})^\top (\mathbf{y} - \mathbf{x}^{(k)}) + \frac{L}{2}\|\mathbf{x}^{(k)} - \mathbf{y}\|^2$$

   (a) Express the iterates, $x^{(k+1)}$, in the form

$$x^{(k+1)} = x^{(k)} + \alpha_k \mathbf{d}^{(k)}$$

   5 marks

   (b) Show that there exists some $i \in \{0, 1, \ldots, N-1\}$ such that

$$\|\nabla f(\mathbf{x}^{(i)})\| \leq \frac{C}{\sqrt{N}}$$

   holds where C is a constant depending on the starting point. 5 marks

5. Consider the following function

$$f(\mathbf{x}) = \mathbf{x}^\top \begin{bmatrix} \frac{1}{2} & -1 \\ 1 & 1 \end{bmatrix} \mathbf{x} + [1\ 2]^\top \mathbf{x} + 10, \mathbf{x} = [x_1\ x_2]^\top$$

(a) Find $\mathbf{x}^*$,the global minimum of f. 5 marks

(b) Consider applying steepest descent procedure with exact line search on this problem. Starting from $\mathbf{x}^{(0)} = [0\ 0]^\top$ compute N, the number of steepest descent iterations, to satisfy

$$f(\mathbf{x}^{(N)}) - f(\mathbf{x}^*) \leq 10^{-3}$$

where $\mathbf{x}^{(N)}$ is the Nth iterate. 5 marks

6. Suppose you are testing an implementation of the*Rank 1 update Quasi Newton Algorithm* to check if it has any bugs. Let $G^{(k)}$ be the kth matrix iterate. It was found that after 5 iterations $G^{(1)}, G^{(3)}$ had all eigenvalues positive but $G^{(2)}, G^{(4)}$ had some negative eigenvalues. Could we infer from these observations that the implementation was buggy. Give reasons 5 marks

$[-1\ 0] \begin{bmatrix} \frac{1}{2} & -1 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \end{bmatrix} \frac{1}{2} + [1\ 2] \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

$\frac{1}{2} - 1$ $\qquad \frac{1}{2} - 1 + 10 = \frac{9}{2}$

$-\frac{1}{2} + 10$

Computational Methods of Optimization

Second ~~First~~ Midterm(4th Nov, 2015)

Answer all questions
Answer all questions Time: 90 minutes

1. Let $Q \succ 0$ be a symmetric matrix. Consider minimizing

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top Q\mathbf{x} - h^\top \mathbf{x}$$

using Quasi-newton procedure by imposing the condition $G^{(k+1)}\gamma^{(k)} = \delta^{(k)}$ where $G^{(k)}$s are chosen from Broyden family, and $\delta^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}, \gamma^{(k)} = \nabla f(\mathbf{x}^{(k+1)}) - \nabla f(\mathbf{x}^{(k)})$, $\mathbf{x}^{(k)}$ is the output of the kth iteration. Show that $\delta^{(k)\top}Q\delta^{(i)} = 0, i \leq k$. 10 marks

2. Let A be a $n \times n$ sqaure symmetric matrix. It is further given that $A = \sum_{i=1}^{j} u_i u_i^\top + 2\sum_{l=1}^{n-j} u_{j+l}u_{j+l}^\top$ where $u_i \in \mathbb{R}^n, \|u_i\| = 1, u_i^\top u_j = 0, i \neq j$. Consider computing $A\mathbf{x} = b$ where b is a vector.

   (a) Describe a conjugate gradient algorithm for solving this problem. 5 marks

   (b) How many iterations will your algorithm take. Justify your answer 5 marks

3. Let $A \succ 0$ be a $n \times n$ symmetric real valued matrix. For a given $\mathbf{y} \in \mathbb{R}^n$, consider the following optimization problem

$$min_{\mathbf{x}} - \mathbf{y}^\top \mathbf{x} \text{ subject to } \mathbf{x}^\top A\mathbf{x} \leq 1$$

   (a) Find the Lagrange dual of the above problem 2 marks

   (b) State and solve the dual optimization problem. 4 marks

   (c) Show that strong duality holds. 1 marks

   (d) Prove the inequality $(\mathbf{y}^\top \mathbf{x})^2 \leq (\mathbf{x}^\top A\mathbf{x})\,(\mathbf{y}^\top A^{-1}\mathbf{y})$ 3 marks

4. Let X be a Random variable with mean m and taking values over positive integers between 1 and n. Let $P(X = i) = p_i$ be the probability of the event $X = i$. We wish to choose $p = [p_1, \ldots, p_n]^\top$ to maximize the entropy. The entropy maximization problem can be formulated as follows

$$min_p \sum_{i=1}^{n} p_i \log p_i \quad \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} ip_i = m, p_i \geq 0, i \in \{1, \ldots, n\}$$

Compute p. 10 marks

$\nabla_x (x^T A x) = 2x^T A.$

5. Let $f(\mathbf{x})$ be defined in Q. 1. Consider the problem

$$min_{\mathbf{x}} f(\mathbf{x}) \ \text{ subject to } \mathbf{a}_i^\top \mathbf{x} \geq b_i \qquad \mathbf{a}_i \in \mathbb{R}^n, b_i \in \mathbb{R}, i \in \{1, \dots, m\}$$

It is given that $\mathbf{x}^0$ is a feasible point

(a) Let $\hat{\mathbf{x}} = \text{argmin}_{\mathbf{x}} f(\mathbf{x})$, subject to $\mathbf{a}_i^\top(\mathbf{x} - \mathbf{x}^0) = 0,\ i \in \{1, \dots, k\}$ .For any $\mathbf{x}(\alpha) = \mathbf{x}^0 + \alpha(\hat{\mathbf{x}} - \mathbf{x}^0)$, $\alpha > 0$, show that $\nabla f(\mathbf{x}^0)^\top(\mathbf{x}(\alpha) - \mathbf{x}^0) < 0$ whenever $\mathbf{x}^0 \neq \hat{\mathbf{x}}$. 5 marks

(b) Suppose $\nabla f(\hat{\mathbf{x}}) = \sum_{j=1}^k \mu_j \mathbf{a}_j$ $\quad k < m$ and $\mu_k < 0$. Let $\tilde{\mathbf{x}}$ be defined as follows

$$\tilde{\mathbf{x}} = \text{argmin}_{\mathbf{x}} f(\mathbf{x}), \ \text{ subject to } \mathbf{a}_i^\top(\mathbf{x} - \hat{\mathbf{x}}) = 0 \ i \in \{1, \dots, k-1\}$$

Show that $\mathbf{a}_k^\top \tilde{\mathbf{x}} \geq b_k$ and $f(\tilde{\mathbf{x}}) < f(\hat{\mathbf{x}})$. 5 marks

Computational Methods of Optimization
Final Exam(1st Dec, 2015)

Answer all questions Time: 180 minutes

1. Let $f : \mathbb{R}^n \to \mathbb{R}$,

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

$Q \succ 0$, real valued symmetric matrix with minimum and maximum eigenvalue b and B respectively

   (a) Let $\mathbf{e}_i$ denote the ith column of $n \times n$ Identity matrix. Doing coordinate descent along the ith coordinate would mean doing gradient descent with descent direction, $\mathbf{d} = s\mathbf{e}_i$ where s is an appropriately chosen scalar. At any point $\mathbf{x}$, compute $\mathbf{d}$, i.e., compute s, if we do coordinate-descent along ith coordinate? 1 mark

   (b) Consider the following descent algorithm.
   At the start of the kth iteration, at the point $\mathbf{x}^{(k)}$, do coordinate descent by the choosing the coordinate according to the rule, $i^* = argmax_i |(\nabla f(\mathbf{x}))_i|$. Let $\mathbf{d}^{(k)}$ be the descent direction. Let $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)}\mathbf{d}^{(k)}$ where $\alpha^{(k)}$ is chosen by exact line search. Show that 5 mark

$$\nabla f(\mathbf{x}^{(k)\top}\mathbf{d}^{(k)} \geq \frac{1}{n}\|\mathbf{d}^{(k)}\|^2\|\nabla f(\mathbf{x}^{(k)})\|^2$$

   (c) Derive the rate of convergence for this algorithm. Compare it with Steepest Descent procedure 4 mark

2. Consider the following problem

$$min_{\mathbf{x}\leq\mathbb{R}^n}\mathbf{x}^\top A\mathbf{x} \quad \|\mathbf{x}\|^2 \leq 1$$

where A is a symmetric $n \times n$ matrix

   (a) Deduce the dual optimization problem for this formulation 4 marks

   (b) Show that strong duality holds and solve the problem 4 marks

   (c) From your answer derive μ_1, μ_2 so that

$$\mu_1 \geq \frac{\mathbf{x}^\top A\mathbf{x}}{\mathbf{x}^\top\mathbf{x}} \geq \mu_2$$

2 marks

3. Consider the following

$$min_x f(\mathbf{x}) \text{ subject to } \|\mathbf{x}\| \leq 1$$

$$f(\mathbf{x}) = x_1^2 - x_2^2 - \frac{18}{5}x_1 + \frac{16}{5}x_2$$

(a) Find a KKT point, $\mathbf{x}^0$. 4 marks

(b) State the Dual problem. 4 marks

(c) The point $\mathbf{x}^0$ is the global minimum of the original problem. Prove or disprove 3 marks

4. Let c be a positive constant. Consider the problem

$$min_{\mathbf{x}\in\mathbb{R}^3} 2e^{x_1} + \|\mathbf{x}\|^2 - 2x_1x_2 - 2(x_2 - x_1)$$

$$\text{subject to } x_1x_2x_3 + x_2x_3 \leq 2,\ x_1 + x_3 \geq c,\ x_1 \geq -1,\ x_2 \geq 0, x_3 \geq 0$$

We examine $\mathbf{x}^0 = [0, 1, 1]^\top$ as a potential candidate solution.

(a) Can you choose c so that $\mathbf{x}^0$ satisfies the first order neccessary conditions. 5 marks

(b) Is there a choice of c such that $\mathbf{x}^0$ is also a global minimum. Justify 5 marks

5. Let $\mathbf{a}, \mathbf{c} \in \mathbb{R}^n$ be non-negative. Consider the following optimization problem

$$min_{\mathbf{x}\geq 0} \sum_{i=1}^{n} \frac{c_i}{x_i} \text{ subject to } \mathbf{a}^\top \mathbf{x} \leq b$$

(a) Find the global optimal point. Justify your answer 7 marks

(b) Using your answer prove that 3 marks

$$\|\mathbf{z}\|^2 \leq min_{\mathbf{x}} \sum_{i=1}^{n} \frac{z_i^2}{x_i} \text{ subject to } \sum_{i=1}^{n} x_i = 1, \mathbf{x} \geq 0$$

6. Let $C = \{x | x \in \mathbb{R}^d,\ a_i \leq x_1 \leq b_i,\ i = 1, \ldots, d\}$. Compute $P_C(z)$ for $z \notin C$. 10 marks

7. Consider minimizing

$$f(\mathbf{x}) = x_1^2 + x_2^2 + x_1x_2 \text{ subject to } x_1 + x_2 = 1$$

by penalty function method.

(a) For any penalty parameter c compute the eigenvalues of the Hessian? 3 marks

(b) Comment on the convergence of the method 2 marks

(c) Consider the augmented Lagrangian method. Recall the function $L_c(\mathbf{x}, \mu)$. Compute $\phi(\mu) = min_{\mathbf{x}} L_c(\mathbf{x}, \mu)$. What is the Hessian of ϕ. Can you contrast the convergence of Augmented Lagrangian and the Penalty function method. 5 marks